Acta Mathematicae Applicatae Sinica, English Series
Vol. 28, No. 4 (2012) 807–822
DOI: 10.1007/s10255-012-0190-2
http://www.ApplMath.com.cn & www.SpringerLink.com

Acta Mathematicae Applicatae Sinica, English Series

Heavy Traffic Limit Theorems for a Queue with Poisson ON/OFF Long-range Dependent Sources and General Service Time Distribution

Wan-yang DAI

Department of Mathematics, and State Key Laboratory of Novel Software Technology, Nanjing University, Nanjing 210093, China (E-mail: nan5lu8@netra.nju.edu.cn)

Abstract In Internet environment, traffic flow to a link is typically modeled by superposition of ON/OFF based sources. During each ON-period for a particular source, packets arrive according to a Poisson process and packet sizes (hence service times) can be generally distributed. In this paper, we establish heavy traffic limit theorems to provide suitable approximations for the system under first-in first-out (FIFO) and work-conserving service discipline, which state that, when the lengths of both ON- and OFF-periods are lightly tailed, the sequences of the scaled queue length and workload processes converge weakly to short-range dependent reflecting Gaussian processes, and when the lengths of ON- and/or OFF-periods are heavily tailed with infinite variance, the sequences converge weakly to either reflecting fractional Brownian motions (FBMs) or certain type of long-range dependent reflecting Gaussian processes depending on the choice of scaling as the number of superposed sources tends to infinity. Moreover, the sequences exhibit a state space collapse-*like* property when the number of sources is large enough, which is a kind of extension of the well-known Little's law for M/M/1 queueing system. Theory to justify the approximations is based on appropriate heavy traffic conditions which essentially mean that the service rate closely approaches the arrival rate when the number of input sources tends to infinity.

Keywords reflecting fractional Brownian motion, reflecting Gaussian process, long-range dependence, queueing process, weak convergence

2000 MR Subject Classification 60F17; 60K25; 90B20; 90B22

1 Introduction

ON/OFF sources are widely used to model voice, video and data traffics in telecommunication systems (see, e.g., [16, 19, 23, 26, 28]). In particular, stochastic modeling of queueing systems with ON/OFF long-range dependent data has become an active area of research. In contrast to most of the existing achievements in this field, which are based on fluid models whose outputs are deterministic with constant (service) rates and whose inputs are certain types of long-range dependent fluid sources (e.g. [10, 11, 21, 24]), we will model our queueing system with general service time distribution and the input as a superposition of Poisson ON/OFF point processes to better capture the variation of packet sizes and the behavior of real packet traffic. Concretely, for a particular source, packets arrive according to a Poisson process during each ON-period. For such a source, the corresponding traffic exhibits long-range dependence (see, for instance, [27]) when the lengths of ON- and/or OFF-periods are heavily tailed with infinite variance. Besides the assumption on the service time distribution, our system is further supposed to operate under FIFO and work-conserving discipline.

A special case of the above queueing model is discussed in [3], where the distributions of

Manuscript received June 20, 2008. Revised April 27, 2011.
Supported by the National Natural Science Foundation of China (No. 10371053, 10971249).

ON- and OFF-periods are assumed to be Pareto and exponential respectively and packet sizes are supposed to be constant. They show that the sequence of probabilities that steady state unfinished works exceed a threshold tend to the corresponding probability assuming Poisson input process when the number of input sources tends to infinity. Currently, it is not clear whether their result can be extended to the more general model as presented above. Furthermore, the dependence of the convergence rates on various parameters of the system is not shown in their result, e.g., the relationship between ρ^N (traffic intensity, utilization level) and N (the number of sources).

Due to the above reasons, we will study our queueing system by employing some other method. Under heavy traffic conditions (suitable relationships between ρ and N such that the service rate closely approaches the arrival rate when N tends to infinity), we will show that, when the lengths of both ON- and OFF-periods are lightly tailed, the sequences of the scaled queue length and workload processes converge weakly to short-range dependent reflecting Gaussian processes, and when the lengths of ON- and/or OFF-periods are heavily tailed with infinite variance, the sequences converge weakly to either reflecting fractional Brownian motions (FBMs) or certain type of long-range dependent reflecting Gaussian processes depending on the choice of scaling as the number of input sources tends to infinity. Moreover, the sequences exhibit a state space collapse-*like* property when N is large enough, which is a kind of extension of the well-known Little's law for M/M/1 queueing system.

Our heavy traffic limits set up certain connection between the above physical queueing systems and some existing fluid queueing models. For example, for a fluid model with constant output rate and FBM input, the stationary queue content distribution is asymptotically Weibullian (e.g., [15, 21, 24], and more generally, as summarized in [29]), namely, the probability of exceeding buffer level b is roughly of the form $\exp(-b^{2(1-H)})$ if FBM is characterized by Hurst parameter H. The result can be applied to derive corresponding probability for our reflecting FBM after properly managing parameters.

Concerning heavy traffic limit theorems for queueing systems with long-range dependent inputs, there are only a few achievements until now besides the one mentioned above in [3]. In [10] and [11], authors studied a fluid queueing system with constant output rate and a superposition of ON/OFF fluid input sources. In [18] and [20], instead of discussing superposition problem, authors considered a single class and feedforward multiclass queueing networks with long-range dependent interarrival and service time sequences respectively. The current limit theorems are the supplements of these existing results. In justifying our reflecting FBM approximation, we will adopt the simultaneous limit regime related to FBM in [22], in which both N (the number of sources) and T (the time-scaling parameter) go to infinity at the same time. This procedure provides us some convenience in employing some ingredient developed in [22] to establish the weak convergence for our scaled queue length and workload processes.

One last point we wish to mention is that we have employed our theorem on reflecting Gaussian processes in the current paper to provide a reasonable interpretation (in [9]) to some well-known large-scale computer and statistical experiments conducted by Cao et al[2], Cao and Ramanan[3], which reveal some gap between their simulation findings and the existing theory on heavy-tail and long range dependence. In [9], the author finds out that all the 'heavy-tail' random variables used in computer and network simulations are truncated versions of their real heavy-tail counterparts due to the limitations of computer hardware and softwares, and hence they are not heavily tailed ones. So, by combining the findings in [9] and the theorem in the current paper, we claim in [9] that the findings in [2, 3] are more close to practice but not to the mathematical assumptions imposed in their models since their simulations are computer-based ones.

To be convenient for readers, here we summarize some frequently used notations and terminologies throughout the paper. First, we recall the definition of u.o.c. convergence. For a function $f : [0, \infty) \to R$ and $t \geq 0$, put

$$\|f\|_t \equiv \sup_{0 \leq s \leq t} |f(s)|,$$

then a sequence of functions $f^n : [0, \infty) \to R$ is said to converge uniformly on compact sets (u.o.c.) to f if for each $t \geq 0$, $\|f^n - f\|_t \to 0$ as $n \to \infty$. Second, we use $C_b(R)$ to denote the set of all bounded and continuous functions f and $C(R)$ to denote the set of all continuous functions over the real number space R, which are endowed with the uniform topology. Third, we use $D_E[0, \infty)$ to denote the Skorohod topological space, i.e., the space of E-valued functions that are right continuous and have left-hand limits, which is endowed with the Skorohod topology (see, e.g., [1, 12]). Fourth, we use i.i.d to denote independent and identically distributed, use a.s. to denote almost surely, use '$\Rightarrow$' to denote 'converge in distribution' or equivalently 'converge weakly', and use '$\sim$' to denote 'equals approximately'.

The rest of this paper is organized as follows. In Section 2, we formulate our model, and in Section 3, we present our main theorems and they are proved in Section 4.

2 Queueing Model Formulation

In this section, we consider a queueing system with general service time distribution and with N i.i.d. Poisson ON/OFF input sources. Concretely, a Poisson ON/OFF source $n \in \{1, .., N\}$ consists of independent strictly alternating ON- and OFF-periods, moreover, it transmits packets to a server according to a Poisson process with interarrival time sequence $\{u_n(i), i \geq 1\}$ and rate λ if it is ON and remains silent if it is OFF. The lengths of the ON-periods are identically distributed and so are the lengths of OFF-periods, and furthermore, both of their distributions can be heavily tailed with infinite variance. Specifically, for any distribution F, we denote by $\overline{F} = 1 - F$ the complementary (or right tail) distribution, and by F_1 and F_2 the distributions for ON- and OFF-periods with probability density functions f_1 and f_2 respectively. Their means and variances are denoted by μ_i and σ_i^2 for $i = 1, 2$. In what follows, we assume that as $x \to \infty$,

$$\text{either } \overline{F}_i(x) \sim x^{-\alpha_i} L_i(x) \text{ with } 1 < \alpha_i < 2 \text{ or } \sigma_i^2 < \infty, \tag{2.1}$$

where $L_i > 0$ is a slowly varying function at infinity, that is,

$$\lim_{x \to \infty} \frac{L_i(tx)}{L_i(x)} = 1 \qquad \text{for any } t > 0.$$

Note that the mean μ_i is always finite but the variance σ_i^2 is infinite when $\alpha_i < 2$, and furthermore, one distribution may have finite variance and the other has an infinite variance since F_1 and F_2 are allowed to be different. The sizes of transmitted packets (service times) form an i.i.d. random sequence $\{v^N(i) = v(i)/\mu^N, i \geq 1\}$, where μ^N is the rate of transmission corresponding to each N and $\{v(i) : i \geq 1\}$ is an i.i.d. random sequence with mean 1 and variance σ_v^2, moreover, $\{v(i) : i \geq 1\}$ is independent of the arrival processes.

To derive our queueing dynamical equation, we introduce more notations. For a single source $n \in \{1, \cdots, N\}$, it follows from the explanation in [22] that the alternating ON/OFF periods can be described by a stationary binary process $W_n = \{W_n(t), t \geq 0\}$: $W_n(t) = 1$ means that input traffic is in an ON-period at time t and $W_n(t) = 0$ means that input traffic is in an OFF-period, and moreover, the mean of W_n is given by

$$\gamma = EW_n(t) = P(W_n(t) = 1) = \mu_1/(\mu_1 + \mu_2). \tag{2.2}$$

Let $T_n(t)$ denote the cumulative amount of time which the nth source is ON during time interval $[0, t]$, that is,

$$T_n(t) = \int_0^t W_n(s) ds. \tag{2.3}$$

Let $A_n(t)$ be the total number of packets arrived at the server from the nth source during $[0,t]$, namely,

$$A_n(t) = \sup\Big\{m, \sum_{i=1}^{m} u_n(i) \le T_n(t)\Big\}, \tag{2.4}$$

which exhibits *long range dependence* if σ_1 and σ_2 are not finite simultaneously (see, for instance, [27]). Moreover, let $A^N(t)$ be the total number of packets transmitted to the server by time t summed over all N sources, that is,

$$A^N(t) = \sum_{n=1}^{N} A_n(t), \tag{2.5}$$

and let $S^N(t)$ be the total number of packets that finished service at the server if her keep busy in $[0,t]$, that is,

$$S^N(t) = \sup\{m, V^N(m) \le t\}, \tag{2.6}$$

where

$$V^N(m) = \sum_{i=1}^{m} v^N(i). \tag{2.7}$$

Then the queue length process $Q^N(t)$ which is the number of packets including the one being served at the server at time t can be represented by

$$Q^N(t) = A^N(t) - S^N(B^N(t)), \tag{2.8}$$

where we assume that the initial queue length is zero for convenience, $B^N(t)$ is the cumulative amount of time that the server is busy by time t. In the following analysis, we will employ FIFO and *non-idling* service discipline under which the server is never idle when there are packets waiting to be served. Hence the total busy time can be represented as

$$B^N(t) = \int_0^t I\{Q^N(s) > 0\}ds,$$

where $I\{\cdot\}$ is the indicator function. Finally, we introduce the below workload process which measures the delay of a packet staying in the system,

$$L^N(t) = V^N(A^N(t)) - B^N(t). \tag{2.9}$$

3 Heavy Traffic Limit Theorems

We are interested in the behaviors of the queueing process $Q^N(\cdot)$ and the workload process $L^N(\cdot)$ under suitable scaling and under the condition that the load of the server closely approaches the service capacity when the source number N gets large enough. In order to state our main theorems, we introduce the below notations for convenience, which are adapted from [28]. When $1 < \alpha_i < 2$, set $a_i = (\Gamma(2-\alpha_i))/(\alpha_i - 1)$. When $\sigma_i^2 < \infty$, set $\alpha_i = 2$, $L_i \equiv 1$ and $a_i = \sigma_i^2/2$. Moreover, let

$$b = \lim_{x\to\infty} t^{\alpha_2-\alpha_1}\frac{L_1(x)}{L_2(x)}.$$

If $0 < b < \infty$ (implying $\alpha_1 = \alpha_2$ and $b = \lim_{x\to\infty} L_1(x)/L_2(x)$), set $\alpha_{\min} = \alpha_1$,

$$\pi^2 = \frac{2(\mu_2^2 a_1 b + \mu_1^2 a_2)}{(\mu_1+\mu_2)^3\Gamma(4-\alpha_{\min})} \qquad \text{and} \qquad L = L_2; \tag{3.1}$$

if, on the other hand, $b = 0$ or $b = \infty$,

$$\pi^2 = \frac{2\mu_{\max}^2 a_{\min}}{(\mu_1+\mu_2)^3\Gamma(4-\alpha_{\min})} \qquad \text{and} \qquad L = L_{\min}, \tag{3.2}$$

where min is the index 1 if $b = \infty$ (e.g. if $\alpha_1 < \alpha_2$) and is the index 2 if $b = 0$, max denoting the other index.

3.1 Reflecting Gaussian Process as the Limit

Condition 3.1 (heavy traffic condition). For each N, let the service rate μ^N be given by

$$\mu^N = N\lambda\gamma + \theta\sqrt{N}, \tag{3.3}$$

where θ is some positive constant.

In addition, we need the below conditions on the distributions of F_1 and F_2:

$$F_i(x) \ \ (i=1,2) \ \text{ is absolutely continuous in terms of } \ x; \tag{3.4}$$

$$\text{The density } \ f_i(x) \ \ (i=1,2) \ \text{ of } \ F_i \ \text{ satisfies } \lim_{x\to 0^+} f_i(x) < \infty. \tag{3.5}$$

Before we state our main theorems, we define the scaling processes for each N as follows,

$$\widetilde{Q}^N(\cdot) \equiv \frac{1}{\sqrt{N}} Q^N(\cdot), \qquad \widetilde{L}^N(\cdot) \equiv \frac{\mu^N}{\sqrt{N}} L^N(\cdot). \tag{3.6}$$

Theorem 3.1. *Under conditions (3.3)–(3.5) and as $N \to \infty$, both $\widetilde{Q}^N(\cdot)$ and $\widetilde{L}^N(\cdot)$ converge in distribution under Skorohod topology to a reflecting Gaussian process $\widetilde{Q}(\cdot)$ given by*

$$\widetilde{Q}(\cdot) = \widetilde{A}(\gamma\cdot) + \lambda\widetilde{T}(\cdot) - \widetilde{S}(\lambda\gamma\cdot) - \theta\cdot + \widetilde{I}(\cdot) \geq 0, \tag{3.7}$$

where the three processes $\widetilde{A}(\gamma\cdot)$, $\widetilde{S}(\lambda\gamma\cdot)$ and $\widetilde{T}(\cdot)$ are independent each other, and furthermore, $\widetilde{A}(\gamma\cdot)$ is a Brownian motion with mean zero and variance function $\lambda\gamma\cdot$, $\widetilde{S}(\lambda\gamma\cdot)$ is also a Brownian motion with mean zero and variance function $\lambda\gamma\sigma_v^2\cdot$, $\widetilde{T}(\cdot)$ is a Gaussian process with a.s. continuous sample paths, mean zero and stationary increments, whose covariance and variance functions satisfy

$$\operatorname{Cov}(\widetilde{T}(t), \widetilde{T}(s)) = \frac{1}{2}\big(\operatorname{Var}(\widetilde{T}(t)) + \operatorname{Var}(\widetilde{T}(s)) - \operatorname{Var}(\widetilde{T}(t-s))\big), \tag{3.8}$$

$$\operatorname{Var}(\widetilde{T}(t)) \sim \begin{cases} \pi^2 t^{2H} L(t) & \text{as } t\to\infty \ \text{ for } \ 1<\alpha_{\min}<2, \\ \pi^2 t & \text{as } t\to\infty \ \text{ and } \ \alpha_{\min}=2, \end{cases} \tag{3.9}$$

where H is the Hurst parameter given by $H = (3-\alpha_{\min})/2$. Moreover, $\widetilde{I}(\cdot)$ in (3.7) is a non-decreasing process with $\widetilde{I}(0) = 0$ and satisfies

$$\int_0^\infty \widetilde{Q}(s)\,d\widetilde{I}(s) = 0.$$

Remark 3.1. More discussions about reflected Gaussian processes, readers are referred to [29]. From the theorem, we have the following observations. When $1 < \alpha_{\min} < 2$, we have that $1/2 < H < 1$ which implies that the process $\widetilde{T}(\cdot)$ exhibits long range dependence. When

$\alpha_i = 2$ for $i = 1, 2$, the ON- and OFF-periods both have finite variance and hence we have that $H = 1/2$ and $L = 1$, which imply that $\widetilde{T}(\cdot)$ exhibits short range dependence. Finally, the results given in the theorem can be considered as a kind of extension of Little's formula for M/M/1 queueing model or considered as satisfying certain state space collapse property.

3.2 Reflecting Fractional Brownian Motion as the Limit

In this subsection, we suppose that at least one of σ_i^2 $(i = 1, 2)$ is infinite. To further discussion, we need to introduce another time-scaling parameter R and assume that $N = N(R)$ goes to infinite as $R \to \infty$. Moreover, we assume that N is taken to satisfy the below fast growth condition (and see more discussion in [22])

$$NR\overline{F}_L(R) \to \infty \qquad \text{as } R \to \infty, \tag{3.10}$$

where $\overline{F}_L = \overline{F}_i$ if $L = L_i$ and L is defined in (3.1) and (3.2). Notice that (3.10) implies $NR^{1-\alpha_{\min}}L(R) \to \infty$.

Condition 3.2 (heavy traffic condition). For each N and R, let the service rate μ^R be given by

$$\mu^R = N\lambda\gamma + \theta\big(NR^{1-\alpha_{\min}}L(R)\big)^{1/2}, \tag{3.11}$$

where θ is some positive constant.

Next, let d_R be the normalization sequence given by

$$d_R = (NR^{3-\alpha_{\min}}L(R))^{1/2}, \tag{3.12}$$

and define

$$\widetilde{Q}^R(\cdot) \equiv \frac{1}{d_R}Q^N(R\cdot), \qquad \widetilde{L}^R(\cdot) \equiv \frac{\mu^R}{d_R}L^N(R\cdot). \tag{3.13}$$

Theorem 3.2. *Assuming that conditions (3.10) and (3.11) hold, then as $R \to \infty$, both $\widetilde{Q}^R(\cdot)$ and $\widetilde{L}^R(\cdot)$ converge in distribution under Skorohod topology to a process $\widetilde{Q}_H(\cdot)$ given by*

$$\widetilde{Q}_H(\cdot) = \lambda\pi B_H(\cdot) - \theta\cdot + \widetilde{I}_H(\cdot) \geq 0, \tag{3.14}$$

where $B_H(\cdot)$ is a standard FBM, and $\widetilde{I}_H(\cdot)$ is a non-decreasing process with $\widetilde{I}(0) = 0$ and satisfies

$$\int_0^\infty \widetilde{Q}_H(s)d\widetilde{I}_H(s) = 0.$$

Remark 3.2. Standard FBM is a mean zero Gaussian process with a.s. continuous sample paths and whose covariance structure is as follows

$$\mathrm{Cov}(B_H(t), B_H(s)) = \frac{1}{2}\big(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\big).$$

4 Proof of Main Theorems

Let $T^N(t)$ be the total cumulative amount of ON time summed over all N sources, that is,

$$T^N(t) \equiv \sum_{n=1}^N T_n(t) = \int_0^t W^N(s)ds, \tag{4.1}$$

where $W^N(\cdot)$ is the superposition of $W_n(\cdot)$ for $n = 1, \cdots, N$, that is, for each $t \in [0, \infty)$,

$$W^N(t) = \sum_{n=1}^{N} W_n(t). \tag{4.2}$$

Moreover, let $\overline{A}(t)$ denote the cumulative number of arrival packets to the server during the time interval $[0, t]$, that is,

$$\overline{A}(t) = \sup\Big\{m : \sum_{i=1}^{m} u(i) \le t\Big\}, \tag{4.3}$$

where $\{u(i), i = 1, 2, \cdots\}$ is an exponentially distributed random sequence with mean value $1/\lambda$, which is independent of all processes mentioned before. Then we have the below lemma.

Lemma 4.1. *The stochastic processes $A^N(\cdot)$ in (2.5) and $\overline{A}(T^N(\cdot))$ in (4.3) have the same distribution.*

Proof. To show that $A^N(\cdot)$ and $\overline{A}(T^N(\cdot))$ have the same distribution, it suffices to show that they have the same finite-dimensional distribution for an arbitrary positive integer k and arbitrary numbers $t_1, \cdots, t_k \in [0, \infty)$ according to Proposition 2.2 in [17].

Notice that the process $W^N(\cdot)$ in (4.2) takes values in the set $\mathcal{N} = \{0, 1, \cdots, N\}$ and has the piecewise constant sample paths given by

$$x(t) = \sum_{i=1}^{M} n_{i-1} I\{s_{i-1} \le t < s_i\}, \qquad n_{i-1} \in \mathcal{N}, \quad n_{i-1} \ne n_i, \tag{4.4}$$

where $s_0, s_1, \cdots, s_M$ with $s_0 = 0$ and $s_M = \infty$ is a partition of the interval $[0, \infty)$ and M is a positive integer or infinite. Then we use $D_{\mathcal{N}}[0, \infty)$ to denote the set of all of these functions defined in (4.4). Obviously, it is a subset of the Skorohod topological space $D_E[0, \infty)$. Under the same topology, $D_{\mathcal{N}}[0, \infty)$ becomes a measurable space in its own right when endowed with the Borel σ-field $A \cap \mathcal{B} = \{A \cap B, B \in \mathcal{B}\}$ where $\mathcal{B}$ is the Borel σ-field in $D_E[0, \infty)$ (see, for example, [17]). Then there is a probability distribution $F^N(\cdot)$ on $D_{\mathcal{N}}[0, \infty)$ for the process $W^N(\cdot)$ in (4.2), which is uniquely determined by the length distributions of ON- and OFF-periods and the source number N (here, for our purpose, we will not derive the explicit expression of $F^N(\cdot)$).

Basing on the above observation, we first consider the one-dimensional case. For each $t \ge 0$ and each nonnegative number m, it follows from the independent and stationary increment properties of Poisson process that

$$P\{A^N(t) = m\} = \int_{D_{\mathcal{N}}[0,\infty)} \Big\{\sum_{i=1}^{c} N_{n_i\lambda}(\Delta s_i) = m | W^N(\cdot) = x(\cdot)\Big\} F^N(dx),$$

where $x(\cdot)$ is a sample path as defined in (4.4), $N_{n_i\lambda}(\Delta s_i)$ is the number of arrival packets for the Poisson process with arrival rate $n_i\lambda$ during the time interval $\Delta s_i = \min\{s_i, t\} - \min\{s_{i-1}, t\}$ for $i \in \{1, \cdots, M\}$, and the integer c is given by $c = 1 + \sup\{i : s_i < t\}$. Then by the independent and stationary increment properties again, we have,

$$\begin{aligned} &P\{A^N(t) = m\} \\ &= \int_{D_{\mathcal{N}}[0,\infty)} P\Big\{\sum_{i=0}^{N} N_{i\lambda}(\Delta \overline{s}_i) = m | W^N(\cdot) = x(\cdot)\Big\} F^N(dx) \\ &= \int_{D_{\mathcal{N}}[0,\infty)} P\Big\{\sum_{i=1}^{N} N_{\lambda}(i\Delta \overline{s}_i) = m | W^N(\cdot) = x(\cdot)\Big\} F^N(dx) \end{aligned}$$

$$= \int_{D_{\mathcal{N}}[0,\infty)} P\big\{N_\lambda(\tau^N(t)) = m | W^N(\cdot) = x(\cdot)\big\} F^N(dx)$$
$$= P\{\overline{A}(T^N(t)) = m\},$$

where $\Delta\overline{s}_i$ is the summation of time intervals during which the arrival rate for the associated Poisson process is $i\lambda$, and $\tau^N(t)$ is the total cumulative amount of ON time from all N sources up to time t along the sample path $x(\cdot)$.

Secondly, we consider the two-dimensional case (we will omit the discussion for more higher-dimensional cases since they are similar). For any $t_1, t_2 \in [0, \infty)$ with $t_1 < t_2$, and nonnegative integers m_1 and m_2, it follows from the independent and stationary increment properties and the definition of conditional probability that

$$\begin{aligned}
&P\{A^N(t_1) = m_1, A^N(t_2) = m_2\} \\
&= \int_{D_{\mathcal{N}}[0,\infty)} P\{A^N(t_1) = m_1 | W^N(\cdot) = x(\cdot)\} \\
&\qquad \cdot P\{A^N(t_2 - t_1) = m_2 - m_1 | W^N(\cdot) = x(\cdot)\} F^N(dx) \\
&= \int_{D_{\mathcal{N}}[0,\infty)} P\{N_\lambda(\tau(t_1)) = m_1 | W^N(\cdot) = x(\cdot)\} \\
&\qquad \cdot P\{N_\lambda(\tau(t_2 - t_1)) = m_2 - m_1 | W^N(\cdot) = x(\cdot)\} F^N(dx) \\
&= P\{\overline{A}(T^N(t_1)) = m_1, \overline{A}(T^N(t_2)) = m_2\},
\end{aligned}$$

where $\tau(t_2 - t_1)$ is the total cumulative amount of ON time from all N sources during time interval $[t_1, t_2)$ along the path $x(\cdot)$. Hence we have proved that $A^N(\cdot)$ and $\overline{A}(T^N(\cdot))$ have the same distribution. □

4.1 Proof of Theorem 3.1

First of all, we define some scaled and centered processes. For each $t \geq 0$ and $N \geq 1$, let

$$\widetilde{\overline{A}}^N(t) \equiv \frac{1}{\sqrt{N}}(\overline{A}(Nt) - \lambda N t), \tag{4.5}$$

$$\widetilde{S}^N(t) \equiv \frac{1}{\sqrt{N}}(S^N(t) - \mu^N t), \tag{4.6}$$

$$\widetilde{T}^N(t) \equiv \frac{1}{\sqrt{N}}(T^N(t) - \gamma N t) = \int_0^t \frac{1}{\sqrt{N}} \sum_{n=1}^N (W_n(s) - \gamma N) ds. \tag{4.7}$$

Then we have the following lemma.

Lemma 4.2. *There exist three independent processes $\widetilde{A}(\cdot)$, $\widetilde{S}(\lambda\gamma\cdot)$ and $\widetilde{T}(\cdot)$ such that*

$$(\widetilde{\overline{A}}^N(\cdot), \widetilde{S}^N(\cdot), \widetilde{T}^N(\cdot)) \Rightarrow (\widetilde{A}(\cdot), \widetilde{S}(\lambda\gamma\cdot), \widetilde{T}(\cdot)) \qquad \text{as} \quad N \to \infty, \tag{4.8}$$

where $\widetilde{A}(\cdot)$ is a Brownian motion with mean 0 and variance function $\lambda\cdot$, $\widetilde{S}(\lambda\gamma\cdot)$ is a Brownian motion with mean zero and variance function $\lambda\gamma\sigma_v^2\cdot$, $\widetilde{T}(\cdot)$ is a Gaussian process with stationary increments, mean 0, stationary increments, whose covariance and variance functions are as given in (3.8)–(3.9).

Proof. First of all, it follows from Functional Central Limit Theorem (e.g., [5]) that

$$\widetilde{\overline{A}}^N(\cdot) \equiv \frac{1}{\sqrt{N}}(\overline{A}(N\cdot) - \lambda N\cdot) \Rightarrow \widetilde{A}(\cdot), \tag{4.9}$$

where $\widetilde{A}(\cdot)$ is a Brownian motion with mean zero and variance $\lambda\cdot$.

Secondly, for each $t \geq 0$, we have,

$$\begin{aligned}\widetilde{S}^N(t) =& \frac{1}{\sqrt{N}}(\sup\{k : v^N(1) + \cdots + v^N(k) \leq t\} - \mu^N t)\\ =& \frac{1}{\sqrt{N}}(\sup\{k : v(1) + \cdots + v(k) \leq \mu^N t\} - \mu^N t)\\ =& \frac{1}{\sqrt{N}}(S_1^N(N\mu_1^N t) - N\mu_1^N t),\end{aligned}$$

where in the last equation, μ_1^N is given by

$$\mu_1^N = \lambda\gamma + \frac{\theta}{\sqrt{N}},$$

and $S_1^N(\cdot)$ is the counting process corresponding to the i.i.d. normalized random sequence $\{v(i), i \geq 1\}$ with mean 1 and variance σ_v^2. It is obvious that $\mu_1^N \to \lambda\gamma$ as $N \to \infty$. Then by Functional Central Limit Theorem (e.g., [5]), we have

$$\widetilde{S}^N(\cdot) \Rightarrow \widetilde{S}(\lambda\gamma\cdot) \qquad \text{as} \quad N \to \infty,$$

where $\widetilde{S}(\lambda\gamma\cdot)$ is a Brownian motion with mean zero and variance function $\lambda\gamma\sigma_v^2\cdot$.

Thirdly, it follows from conditions (3.4)–(3.5) and Corollary 3.1 in [11] that the below convergence in distribution is true

$$\widetilde{W}^N(\cdot) \equiv \frac{1}{\sqrt{N}}\sum_{n=1}^{N}(W_n(\cdot) - \gamma N) \Rightarrow \widetilde{W}(\cdot), \tag{4.10}$$

where $\widetilde{W}(\cdot)$ is a stationary centered Gaussian process with a.s. continuous sample paths (by [14] since $W_n(t)$ is stochastically continuous) and covariance function $\eta(\cdot)$ which satisfies (see the proof of Theorem 1 in [28] for details),

$$\mathrm{Var}\Big(\int_0^t \widetilde{W}(u)du\Big) = 2\int_0^t\int_0^v \eta(u)dudv, \tag{4.11}$$

which has the expression as in (3.9). By Skorohod representation theorem (see, for example, [12]), we can assume that the convergence in (4.10) is u.o.c. Then we have

$$\widetilde{T}^N(\cdot) = \int_0^\cdot \widetilde{W}^N(s)ds \to \int_0^\cdot \widetilde{W}(s)ds \equiv \widetilde{T}(\cdot) \qquad \text{u.o.c. as} \quad N \to \infty.$$

Thus by the definition of weak convergence on $C[0,\infty)$ (see, for example, [29]), Skorohod representation theorem and Proposition 14.6 in [17], the above u.o.c. convergence implies weak convergence. Now, we show that $\widetilde{T}(\cdot)$ is a Gaussian process. Due to (4.11) and Theorem 7 in [13, p.128], $\widetilde{W}(\cdot)$ is mean square integrable in any given finite interval $[0,T]$, and therefore it follows from Theorem 3 in [13, p.142] that $\widetilde{T}(\cdot)$ is a Gaussian process in $[0,T]$. Since for any given $n \in \{1, 2, \cdots,\}$ and any given $t_1, \cdots, t_n \in [0,\infty)$, we can find an $T_1 < \infty$ such that $t_1, \cdots, t_n$ belong to the common interval $[0,T_1]$. Hence the joint distribution of $\widetilde{T}(t_1), \cdots, \widetilde{T}(t_n)$ is normal. Thus we can conclude that $\widetilde{T}(\cdot)$ is a Gaussian process in $[0,\infty)$, whose variance function is as shown in (4.11). Since $\widetilde{W}(\cdot)$ is stationary, $\widetilde{T}(\cdot)$ has stationary increments and its covariance function is given by the expression in (3.8) due to Proposition 1(b) in [6].

Finally, by the independence assumptions and definitions of related processes, we know that the three processes $\widetilde{\overline{A}}^N(\cdot)$, $\widetilde{S}^N(\cdot)$ and $\widetilde{T}^N(\cdot)$ are independent each other for each N. Thus we can conclude that $\widetilde{A}(\cdot)$, $\widetilde{S}(\lambda\gamma\cdot)$ and $\widetilde{T}(\cdot)$ are independent each other. Hence we finish the proof of the lemma. □

To complete the proof of the theorem, for each $t \geq 0$, we rewrite (2.8) as the summation of centered processes and regulated non-decreasing process as follows,

$$Q^N(t) = X^N(t) + I^N(t), \tag{4.12}$$

where

$$\begin{aligned}
X^N(t) &= (A^N(t) - \lambda\gamma Nt) - (S^N(B^N(t)) - \mu^N B^N(t)) - \sqrt{N}\theta t,\\
I^N(t) &= \mu^N \int_0^t I\{Q^N(s) = 0\}ds.
\end{aligned}$$

The process $I^N(\cdot)$ is non-decreasing process and can increase only when the queue length process $Q^N(\cdot)$ reaches zero due to the non-idling service discipline and the fact that $Q^N(t) \geq 0$ for all $t \geq 0$.

Lemma 4.3.

$$\widetilde{X}^N(\cdot) \equiv \frac{1}{\sqrt{N}} X^N(\cdot) \Rightarrow \widetilde{X}(\cdot) = \widetilde{A}(\gamma\cdot) + \lambda\widetilde{T}(\cdot) - \widetilde{S}(\lambda\gamma\cdot) - \theta\,\cdot \qquad \textit{as } N \to \infty,$$

where $\widetilde{A}(\gamma\cdot)$ is a Brownian motion with mean 0 and variance function $\lambda\gamma\cdot$.

Proof. First of all, we prove the following claim to be true

$$\widetilde{E}^N(\cdot) \equiv \frac{1}{\sqrt{N}}(A^N(\cdot) - \lambda\gamma N\cdot) \Rightarrow \widetilde{A}(\gamma\cdot) + \lambda\widetilde{T}(\cdot) \qquad \text{as } N \to \infty.$$

In fact, by Lemma 4.1, it suffices to prove the following claim,

$$\frac{1}{\sqrt{N}}(\overline{A}(T^N(\cdot)) - \lambda\gamma N\cdot) \Rightarrow \widetilde{A}(\gamma\cdot) + \lambda\widetilde{T}(\cdot) \qquad \text{as } N \to \infty,$$

and it is a direct conclusion of Lemma 4.2 and Corollary 13.3.2 of [29]. Thus, by Lemma 4.2 and the independence assumption, we have the below joint weak convergence

$$(\widetilde{E}^N(\cdot), \widetilde{S}^N(\cdot)) \Rightarrow (\widetilde{A}(\gamma\cdot) + \lambda\widetilde{T}(\cdot), \widetilde{S}(\lambda\gamma\cdot)). \tag{4.13}$$

Moreover, by Skorohod representation theorem, we can assume that the above convergence is u.o.c. a.s. Thus it follows from (4.13) that

$$\Big(\frac{1}{N}A^N(\cdot), \frac{1}{N}S^N(\cdot)\Big) \to (\lambda\gamma\cdot, \lambda\gamma\cdot) \qquad \text{u.o.c. a.s.} \tag{4.14}$$

Then, due to (4.14), the conditions stated in Theorem 6.5 of [5] are satisfied. So, by the same theorem of [5], we know that, for each $t \geq 0$ and as $N \to \infty$,

$$\max_{0 \leq s \leq t} |B^N(s) - s| \to 0. \tag{4.15}$$

Therefore, by the above discussions and the fact that the associated limiting processes have a.s. continuous sample paths, we have

$$\begin{aligned}
&\left\|\widetilde{X}^N(\cdot)-\widetilde{X}(\cdot)\right\|_t \\
\le&\left\|\widetilde{E}^N(\cdot)-\widetilde{A}(\gamma\cdot)+\lambda\widetilde{T}(\cdot)\right\|_t+\left\|\widetilde{S}^N(B^N(\cdot))-\widetilde{S}(\lambda\gamma B^N(\cdot))\right\|_t \\
&+\left\|\widetilde{S}(B^N(\cdot))-\widetilde{S}(\lambda\gamma\cdot)\right\|_t \\
\le&\left\|\widetilde{E}^N(\cdot)-\widetilde{A}(\gamma\cdot)+\lambda\widetilde{T}(\cdot)\right\|_t+\left\|\widetilde{S}^N(\cdot)-\widetilde{S}(\lambda\gamma\cdot)\right\|_t \\
&+\left\|\widetilde{S}(B^N(\cdot))-\widetilde{S}(\lambda\gamma\cdot)\right\|_t \\
\to&0 \qquad \text{a.s.} \quad \text{as } N\to\infty,
\end{aligned}$$

where in the second inequality, we used the fact that $B^N(t)\le t$ for each $t\ge 0$ and in the last claim, we also used the fact that $\widetilde{S}(\cdot)$ is continuous. Thus

$$\widetilde{X}^N(\cdot)\to\widetilde{X}(\cdot) \qquad \text{u.o.c. a.s. as } N\to\infty.$$

Hence by Proposition 5.3 in Chapter 3 of [12], the lemma is proved. □

Next, similar to the discussion as in (4.12), let

$$\widetilde{V}^N(t)=\frac{1}{\sqrt{N}}(\mu^N V^N(t)-t), \tag{4.16}$$

and rewrite (2.9) as the summation of centered processes and regulated non-decreasing process as follows,

$$\mu^N L^N(t)=Z^N(t)+I^N(t), \tag{4.17}$$

where

$$Z^N(t)=(\mu^N V^N(A^N(t))-A^N(t))+(A^N(t)-N\lambda\gamma t)-\sqrt{N}\theta t.$$

Then we have the following lemma.

Lemma 4.4.

$$(\widetilde{V}^N(\cdot),\widetilde{\widetilde{A}}^N(\cdot),\widetilde{T}^N(\cdot))\Rightarrow(\widetilde{V}(\cdot),\widetilde{A}(\lambda\gamma\cdot),\widetilde{T}(\cdot)) \qquad \text{as} \quad N\to\infty, \tag{4.18}$$

where $\widetilde{V}(\cdot)$, $\widetilde{A}(\cdot)$ and $\widetilde{T}(\cdot)$ are independent Brownian motions, moreover, $\widetilde{V}(\cdot)$ is of mean zero and variance function $\sigma_v^2\cdot$, $\widetilde{A}(\cdot)$ and $\widetilde{T}(\cdot)$ are given as before. Moreover,

$$\widetilde{Z}^N(\cdot)\equiv\frac{1}{\sqrt{N}}Z^N(\cdot)\Rightarrow\widetilde{X}(\cdot)=\widetilde{A}(\gamma\cdot)+\lambda\widetilde{T}(\cdot)-\widetilde{S}(\lambda\gamma\cdot)-\theta\,\cdot \qquad \text{as } N\to\infty. \tag{4.19}$$

Proof. By applying Functional Central Limit Theorem and the same explanation as in Lemma 4.2, one can prove the convergence stated in (4.18). Then it follows from (4.18), Lemma 4.1 and random time change theorem that

$$\widetilde{Z}^N(\cdot)\equiv\frac{1}{\sqrt{N}}Z^N(\cdot)\Rightarrow\widetilde{V}(\lambda\gamma\cdot)+\widetilde{A}(\gamma\cdot)+\lambda\widetilde{T}(\cdot)-\theta\,\cdot \qquad \text{as } N\to\infty.$$

Notice that $\widetilde{V}(\lambda\gamma\cdot)$ and $-\widetilde{S}(\lambda\gamma\cdot)$ have the same distribution, we can conclude that the claim stated in (4.19) is true. □

Proof of Theorem 3.1 Once the above lemmas are obtained, we can go over the following standard procedure to finish the proof of the theorem. By Skorohod representation theorem, we suppose that the convergence in Lemma 4.3 is u.o.c. Then, by (4.12) and according to Theorem 6.1 in [5], there uniquely exist a pair of regulated mappings ϕ and ψ, which are continuous, such that for each $t \geq 0$,

$$\widetilde{I}^N(t) = \frac{1}{\sqrt{N}} I^N(t) = \phi(\widetilde{X}^N(t)) = \sup_{0\leq s\leq t} (\widetilde{X}^N(s))^-,$$
$$\widetilde{Q}^N(t) = \frac{1}{\sqrt{N}} Q^N(t) = \psi(\widetilde{X}^N(t)) = \widetilde{X}^N(t) + \phi(\widetilde{X}^N(t)) \geq 0,$$

where $x^-(s) = \max\{-x(s), 0\}$. Then by continuous mapping theorem and Lemma 4.3, we have, as $N \to \infty$,

$$I^N(\cdot) \to \widetilde{I}(\cdot) \equiv \phi(\widetilde{X}(\cdot)) \qquad \text{a.s.} \quad \text{u.o.c.}, \tag{4.20}$$
$$\widetilde{Q}^N(\cdot) \to \widetilde{Q}(\cdot) \equiv \psi(\widetilde{X}(\cdot)) \geq 0 \qquad \text{a.s.} \quad \text{u.o.c.} \tag{4.21}$$

Obviously, $\widetilde{I}(\cdot)$ and $\widetilde{Q}(\cdot)$ have a.s. continuous sample paths, and moreover, $\widetilde{I}(\cdot)$ is non-decreasing with $\widetilde{I}(0) = 0$. Since $\widetilde{Q}^N(t) \geq 0$ and $I^N(t)$ increases only at times t such that $\widetilde{Q}^N(t) = 0$, we have for each $T > 0$,

$$\int_0^T \widetilde{Q}^N(t) \wedge 1 d\widetilde{I}^N(t) = 0. \tag{4.22}$$

Define

$$f : x \in R \to f(x) = x \wedge 1.$$

Clearly, we have $f \in C_b(R)$. Then by (4.20), (4.21), (4.22) and Lemma 8.3 in [8], we have

$$\int_0^T \widetilde{Q}(t) \wedge 1 d\widetilde{I}(t) = 0 \qquad \text{for all} \quad T > 0.$$

Hence $\widetilde{I}(\cdot)$ increases only at times t such that $\widetilde{Q}(t) = 0$.

Finally, by Lemma 4.4 and the same procedure as above, one can prove the weak convergence for the processes of $\widetilde{L}^N(\cdot)$ as $N \to \infty$. □

4.2 Proof of Theorem 3.2

Lemma 4.5 *Let $\beta = 1 - \alpha_{\min}/2$. Then, as $T \to \infty$, we have,*

$$U(T) \equiv T^\beta L(T)^{1/2} \to \infty, \tag{4.23}$$
$$V(T) \equiv T^{\alpha_{\min}/2-1/2}/L(T)^{1/2} \to \infty. \tag{4.24}$$

Proof. Since $L(T)$ is a slowly varying function and $1 < \alpha_{\min} < 2$, we know that $U(T)$ is a regularly varying function with index $0 < \beta < 1/2$, that is, for $x > 0$,

$$\lim_{T\to\infty} \frac{U(Tx)}{U(T)} = x^\beta.$$

Then, take $0 < \epsilon < \beta$, it follows from Proposition 0.8 in [25] that there is a fixed T_0 such that for $x \geq 1$ and $T \geq T_0$, we have

$$U(Tx) > (1 - \epsilon)x^{\beta-\epsilon}U(T).$$

Let $x \to \infty$ in the above inequality, we know that (4.23) is true.

Similarly, $V(T)$ is a regularly varying function with index $0 < \alpha_{\min}/2 - 1/2 < 1/2$, then by the same reason as above, we know that (4.24) holds. □ □

Now for each $t \geq 0$, we rewrite (3.13) as the summation of centered processes and regulated non-decreasing process as follows,

$$\widetilde{Q}^R(t) = X^R(t) + I^R(t), \tag{4.25}$$

where

$$X^R(t) = \frac{1}{d_R}(A^N(Rt) - \lambda\gamma NRt) - \frac{1}{d_R}(S^N(B^N(Rt)) - \mu^R B^N(Rt)) - \theta t,$$
$$I^R(t) = \frac{\mu^R}{d_R} Y^N(Rt) = \frac{R\mu^R}{d_R}\int_0^t I\{\widetilde{Q}^R(s) = 0\}ds.$$

The process $I^R(\cdot)$ is non-decreasing process and can increase only when the queue length process $\widetilde{Q}^R(\cdot)$ reaches zero due to the non-idling service discipline.

Lemma 4.6. *For each N and R and under conditions (3.10) and (3.11), we have, as $R \to \infty$, $X^R(\cdot)$ converges weakly to a process $\widetilde{X}(\cdot)$, that is,*

$$X^R(\cdot) \Rightarrow \widetilde{X}(\cdot) = \lambda\pi B_H(\cdot) - \theta\cdot \tag{4.26}$$

where π and $B_H(\cdot)$ are given in Theorem 3.2.

Proof. Due to Lemma 4.1, it suffices to prove the below facts, as $R \to \infty$,

$$\begin{aligned}&\frac{1}{d_R}\overline{A}(T^N(R\cdot)) - \frac{1}{d_R}\mu^R R\cdot \\ =&\frac{1}{d_R}(\overline{A}(T^N(R\cdot)) - \lambda T^N(R\cdot)) + \frac{1}{d_R}\lambda(T^N(R\cdot) - \gamma NR\cdot) - \theta\cdot \\ \Rightarrow&\lambda\pi B_H(\cdot) - \theta\cdot,\end{aligned} \tag{4.27}$$

and

$$\frac{1}{d_R}\big(S^N(B^N(R\cdot)) - \mu^R B^N(R\cdot)\big) \Rightarrow 0. \tag{4.28}$$

As a matter of fact, notice that from the proof of Theorem 1 in [28], we know that the process $T_n(\cdot)$ defined in (2.3) has variance

$$\mathrm{Var}(T_n(t)) \sim \pi^2 t^{3-\alpha_{\min}} L(t) \qquad \text{as} \quad t \to \infty.$$

Then it follows from condition (3.10) and a similar proof as used in justifying Theorem 4 in [22] that the below weak convergence in the space $C[0,\infty)$ is true,

$$\widetilde{T}^R(\cdot) \equiv \frac{1}{d_R}(T^N(R\cdot) - \gamma NR\cdot) \Rightarrow \pi B_H(\cdot) \qquad \text{as} \quad R \to \infty, \tag{4.29}$$

where B_H is standard fractional Brownian motion with $H = (3 - \alpha_{\min})/2$ and π is given in (3.1) and (3.2).

Next, by Functional Central Limit Theorem (see, for example, [5]), we have that

$$\widetilde{\overline{A}}^R(\cdot) \equiv \frac{1}{(NR)^{1/2}}(\overline{A}(NR\cdot) - \lambda NR\cdot) \Rightarrow \xi^a(\cdot) \qquad \text{as} \ \ R \to \infty, \tag{4.30}$$

where the weak convergence is in the Skorohod topology and $\xi^a(\cdot)$ is a Brownian motion with mean zero and variance λ.

Moreover, for each $t \geq 0$, we have,

$$
\begin{aligned}
\widetilde{S}^R(t) \equiv & \frac{1}{(NR)^{1/2}}(S^N(Rt) - \mu^R Rt) \\
= & \frac{1}{(NR)^{1/2}}(\sup\{k : v(1) + \cdots + v(k) \leq \mu^R Rt\} - \mu^R Rt) \\
= & \frac{1}{(NR)^{1/2}}(S_1^N(NR\mu_1^R t) - NR\mu_1^R t),
\end{aligned}
$$

where in the last equation, μ_1^R is given by

$$\mu_1^R = \lambda\gamma + (R^{1-\alpha_{\min}} L(R))^{1/2}\theta,$$

and $S_1^N(\cdot)$ is the counting process corresponding to the i.i.d. normalized random sequence $\{v(i), i \geq 1\}$ with mean 1. Moreover, by Lemma 4.5, we have that

$$\mu_1^R \to \lambda\gamma \qquad \text{as} \quad R \to \infty.$$

Then by Functional Central Limit Theorem, we have

$$\widetilde{S}^R(\cdot) \Rightarrow \xi^s(\lambda\gamma\cdot) \qquad \text{as} \quad R \to \infty, \tag{4.31}$$

where $\xi^s(\lambda\gamma\cdot)$ is a Brownian motion with mean zero and variance $\lambda\gamma\sigma_v^2$.

Now notice the independent assumption among the processes $\widetilde{\overline{A}}^R(\cdot)$, $\widetilde{S}^R(\cdot)$ and $\widetilde{T}^R(\cdot)$, and the properties that Brownian motion and fractional Brownian motion have a.s. continuous sample paths, then by Skorohod representation theorem (see, for example, [12]), we can and will assume that the convergence in (4.29)–(4.31) is u.o.c. Thus, by Lemma 4.5 and for each $t \geq 0$, as $R \to \infty$,

$$
\begin{aligned}
& \left\| \frac{1}{NR}(T^N(R\cdot) - \gamma NR\cdot) \right\|_t \\
= & \frac{1}{N^{1/2}V(R)} \left\| \frac{1}{d_R}(T^N(R\cdot) - \gamma NR\cdot) \right\|_t \to 0 \qquad \text{a.s.},
\end{aligned}
$$

which implies that as $R \to \infty$,

$$T_1^R(\cdot) \equiv \frac{1}{NR} T^N(R\cdot) \to \gamma\, \cdot \qquad \text{a.s.} \quad \text{u.o.c.} \tag{4.32}$$

Therefore, by (4.30), (4.32), Random Change of Time Theorem in [1] and Lemma 4.5, we have

$$
\begin{aligned}
& \frac{1}{d_R}(\overline{A}(T^N(R\cdot)) - \lambda T^N(R\cdot)) \\
= & \frac{1}{U(R)(NR)^{1/2}}(\overline{A}^N(NRT_1^R(\cdot)) - \lambda NRT_1^R(\cdot)) \to 0 \quad \text{a.s.} \quad \text{u.o.c.}
\end{aligned} \tag{4.33}
$$

Next, notice that, for each $t \geq 0$,

$$B_1^R(t) \equiv \frac{B^N(Rt)}{R} \leq t.$$

Then, it follows from (4.31) and Lemma 4.5 that

$$\begin{aligned}
&\left\| \frac{1}{d_R}(S^N(B^N(Rt)) - \mu^R B^N(Rt)) \right\|_t \\
=&\frac{1}{U(R)} \left\| \frac{1}{(NR)^{1/2}}(S^N(RB_1^N(\cdot)) - \mu^R RB_1^N(\cdot)) \right\|_t \\
\le&\frac{1}{U(R)} \left\| \frac{1}{(NR)^{1/2}}(S^N(R\cdot) - \mu^R R\cdot) \right\|_t \to 0 \qquad \text{a.s.}
\end{aligned}$$

Thus, we have, as $R \to \infty$,

$$\frac{1}{d_R}(S^N(\overline{B}^N(R\cdot)) - \mu^R \overline{B}^N(R\cdot)) \to 0 \qquad \text{a.s.} \quad \text{u.o.c.} \tag{4.34}$$

Hence by (4.29), (4.33) and (4.34), as $R \to \infty$, the convergence stated in (4.27) and (4.28) is true. □

The remaining proof of Theorem 3.2 is similar to that used in justifying Theorem 3.1. Hence we omit it here.

References

[1] Billingsley, P. Convergence of Probability measures. Wiley & Sons, New York, 1968

[2] Cao, J., Cleverland, W.S., Lin, D., Sun, D.X. The effect of statistical multiplexing on the long-range dependence of Internet packet traffic. Tech Report, Bell Labs, Murray Hill, NJ, U.S.A., 2002

[3] Cao J., Ramanan, K. A Poisson limit for buffer overflow probabilities. *Proceedings of Infocom 2002*, 2: 994–1003 (2002)

[4] Cavanaugh, J.D., Salo, T.J. Internetwotking with ATM WANS. In: Advances in Local and Metropolitan Area Networks, ed. by William Stallings, IEEE Computer Society Press, 1994

[5] Chen, H., Yao, D.D. Fundamentals of Queueing Networks: Performance, Asymptotics, and Optimization. Springer-Verlag, New York, 2001

[6] Choe, J., Shroff, N.B. On the supremum distribution of integrated stationary Gaussian processes with negative linear drift. *Adv. Appl. Prob.*, 31: 135–157 (1999)

[7] Chung, K.L., Williams, R.J. Introduction to Stochastic Integration, Birkhauser, Boston, 1983

[8] Dai, J.G., Dai, W. A heavy traffic limit theorem for a class of open queueing networks with finite buffers. *Queueing Systems*, 32: 5–40 (1999)

[9] Dai, W. On the conflict of truncated random variable vs. heavy-tail and long range dependence in computer and network simulation. *Journal of Computational Information System*, 7(5): 1488–1499 (2011)

[10] Debicki, K., Mandjes, M. Traffic with an fBm limit: Convergence of the stationary workload process. *Queueing Systems*, 46: 113–127 (2004)

[11] Debicki, K., Palmowski, Z. On-off fluid models in heavy traffic environment. *Queueing Systems*, 33: 327–338 (1999)

[12] Ethier, S.N., Kurtz, T.G. Markov Processes: Characterization and Convergence, Wiley, New York, 1986

[13] Fudan Uni. Probability Theory: Stochastic Processes. People's Educational Press, 1981

[14] Hahn, M.J. Central limit theorems in $D[0,1]$. *Wahrscheinlichkeitstheorie Verw, Gebiete*, 44, 89–101 (1978)

[15] Husler, J., Piterbarg, V. Extremes of a certain class of Gaussian processes. *Stochastic Processes and Their Applications*, 83: 257–271 (1999)

[16] Jain, R., Routhier, S.A. Packet trains: measurements and a new model for computer netwoek traffic. *IEEE Journal on Selected Areas in Communications*, 4: 986–995 (1986)

[17] Kallenberg, O. Foundations of modern probability. Springer-Verlag, Berlin, 1997

[18] Konstantopoulos, T., Lin, S.J. Fractional Brownian approximations of queueing networks. In: Stochastic Networks: Stability and Rare Events, ed. by P. Glasserman, K. Sigman, K. and D. Yao, Lecture Notes in Statistics, 117, 257–274, Springer-Verlag, New York, 1996

[19] Leland, W.E, Taqqu, M.S., Willinger, W., Wilson, D.V. On the self-similar nature of Ethernet traffic. *IEEE/ACM Trans. Netw.*, 2: 1–15 (1994)

[20] Majewski, K. Fractional Brownian heavy traffic approximations of multiclass feedforward queueing networks. *Queueing Systems*, 50: 199–230 (2005)

[21] Massoulie, L., A. Simonian, A. Large buffer asymptotics for the queue with fractional Brownian input. *J. Appl. Prob.*, 36: 894–906 (1999)

[22] Mikosch, T., Resnick, S., Rootzen, H., Stegeman, A. Is network traffic approximated by stable Levy motion or fractional Brownian motion? *Annals of Applied Probability*, 12: 23–68 (2002)

[23] Nikolaidis, I., Akyildiz, I.F. An overview of source characterization in ATM networks. In: Modeling and Simulation of Computer and Communication Networks: Techniques, Tools and Tutorials, Gordon & Breach Publishing Co, 123–150, 1997

[24] Norros, I. A storage model with self-similar input. *Queueing Systems*, 16: 387–396 (1994)

[25] Resnick, S.I. Extreme Values, Regular Variation and Point Processes. Springer-Verlag, New York, 1987

[26] Paxson, V., Floyd, S. Wide-area traffic: the failure of Poisson modeling. *IEEE/ACM Transactions on Networking*, 3: 226–244 (1995)

[27] Ryu, B.K., Lowen, S.B. Point process approaches to the modeling and analysis of self-similar traffic: Part I-Model construction. *Proc. IEEE Infocom*, 1996: 1468–1475 (1996)

[28] Taqqu, M.S., Willinger, W., Sherman, R. Proof of a fundamental result in self-similar traffic modeling. *ACM/Sigcomm Computer Communication Review*, 27: 5–23 (1997)

[29] Whitt, W. Stochastic Process Limits. Springer-Verlag, New York, 2002

[30] Willinger, W., Taqqu, M.S., Sherman, R., Wilson, D.V. Self-similarity through high-variability: statistical analysis of Ethernet LAN traffic at the source level. *IEEE/ACM Transactions on Networking*, 5(1): 71–86 (2007)