# Distributional Little's law for queues with heterogeneous server interruptions

P. Gao, S. Wittevrongel, K. Laevens, D. De Vleeschauwer and H. Bruneel

**Abstract**

Distributional forms of Little's law relate the steady-state distributions of the number of customers in a queueing system (system content) and the time a customer spends in the system (delay). We discuss a new law for discrete-time multiserver queues with single-slot service times, a first-come-first-served (FCFS) discipline and heterogeneous server interruptions.

*Introduction:* Discrete-time queueing models are studied intensively because of their suitability to describe congestion phenomena in digital communication systems. An important trend in the queueing theory literature is the development of laws that connect the system content and the customer delay. The most well-known result is Little's law, which is valid for any arrival process, service process or scheduling discipline, but only deals with the first moments of system content and delay. Distributional forms of Little's law relate their distributions, see e.g. [1]- [5], in each case however for a specific class of queueing systems. In this letter, we derive a new distributional law for multiserver queues with heterogeneous server interruptions. The law is quite *general* in the sense that it doesn't depend explicitly on the customer arrival process (be it independent or correlated from slot to slot). It is an extension of [3] (no server interruptions) and [4] (homogeneous interruptions and independent arrivals). Server interruptions naturally occur in many applications due to e.g. the breakdown of communication channels, machine repair or processor failures. Therefore we believe that our law is a powerful tool for performance evaluation purposes.

*System description:* We consider a discrete-time multiserver queueing system with infinite buffer size. Time is divided into fixed-length slots. Customers (or "packets") arrive to the system according to a general arrival process and are served (or "transmitted") by an available server (or "output channel") in a FCFS manner. The transmission of a packet can start or end at slot boundaries only and takes exactly one slot. There are $c$ groups of servers, where group $r$ contains $c_r$ servers. Each group is subject to server interruptions independently from group

to group. The number of available servers in group $r$ during slot $k$ is denoted by $t_{r,k}$, and for different $k$-values the $t_{r,k}$'s are independent and identically distributed (i.i.d.) variables, with common probability generating function (pgf) $T_r(z) \triangleq E[z^{t_{r,k}}]$. The service and arrival processes are mutually independent. Finally, the system is assumed to reach a steady state.

Let $v_k$ be the system content (including packets under transmission, if any) at the start of slot $k$, and $a_k$ the number of packet arrivals in slot $k$. Then the following system equation holds:

$$v_{k+1} = \max\left(0, v_k - \sum_{r=1}^{c} t_{r,k}\right) + a_k \,. \tag{1}$$

Also, we define the steady-state joint probability

$$b(i,j) \triangleq \text{Prob}[v = i, a = j] = \lim_{k\to\infty} \text{Prob}[v_k = i, a_k = j] \,, \tag{2}$$

with corresponding joint pgf $B(z,x) \triangleq E[z^v x^a]$.

*Distributional law:* We define the delay of a packet as the total number of slots between the end of the packet's arrival slot and the end of the slot where the packet is transmitted. In this letter, we prove the following relationship between the steady-state pgf $V(z)$ of the system content at the start of an arbitrary slot and the steady-state pgf $D(z)$ of the delay of an arbitrary packet:

$$D(z) = \frac{z-1}{\lambda z} \sum_{p=0}^{C-1} \frac{-1}{T'(x_p)\,(1-x_p)^2} \left\{ \frac{1-z}{z} V\left(\frac{1}{x_p}\right) + \sum_{m_1=0}^{c_1} \cdots \sum_{m_c=0}^{c_c} \left(\prod_{r=1}^{c} t_r(m_r)\right) \sum_{i=0}^{m-1} \left(1 - {x_p}^{m-i}\right) v(i) \right\}, \tag{3}$$

where $\lambda$ is the mean number of arrivals per slot, $C$ is the number of servers, $m = \sum_{r=1}^{c} m_r$, $t_r(m_r) = \text{Prob}[t_r = m_r]$, $v(i) = \text{Prob}[v = i]$, $T(x) = \prod_{r=1}^{c} T_r(x)$, the $x_p$'s are the $C$ solutions for $x$ in terms of $z$ of $1 - z\,T(x) = 0$, which we assume distinct, and $T'(x_p)$ is the first derivative of $T(x)$ with respect to $x$ at $x = x_p$.

*Proof:* Let us consider an arbitrary packet P, that arrives in the queueing system during some slot $J$ in the steady state. Let $d$ with pgf $D(z)$ be the delay of P. To derive $D(z)$, we first make the following observations. The delay of P depends on the number of packets $q$ in the system right after slot $J$ with service priority over P. As long as at the start of slot $J+i$ there are at least $\sum_{r=1}^{c} t_{r,J+i}$ packets in the system with service priority over P, the available servers are all busy

serving packets and P is still waiting for service in slot $J+i$, and there are $\sum_{r=1}^{c} t_{r,J+i}$ departures at the end of slot $J+i$. We may therefore conclude that

$$d > i \quad \Longleftrightarrow \quad q \geq s_i \,, \tag{4}$$

where the random variables $s_i$ are defined as

$$s_0 \triangleq 0; \quad s_i \triangleq \sum_{n=1}^{i} \sum_{r=1}^{c} t_{r,J+n} \,, \quad i \geq 1 \,. \tag{5}$$

Note that as the variables $t_{r,J+i}$ in (5) are i.i.d. from slot to slot and independent from group to group, the pgf of $s_i$ equals

$$S_i(z) \triangleq E[z^{s_i}] = \prod_{r=1}^{c} T_r(z)^i = T(z)^i \,. \tag{6}$$

Secondly, we transform (4) into a relationship between $D(z)$ and the pgf $Q(z)$ of $q$. Using the independence of $q$ and the $s_i$'s and the probability generating property of pgfs, we get

$$\begin{aligned} \frac{D(z)-1}{z-1} &= \sum_{i=0}^{\infty} z^i \sum_{j=0}^{\infty} \text{Prob}[q=j] \, \text{Prob}[q \geq s_i \,|\, q=j] \\ &= \sum_{j=0}^{\infty} \text{Prob}[q=j] \sum_{n=0}^{j} \sum_{i=0}^{\infty} \frac{1}{n!} \, \frac{\mathrm{d}^n}{\mathrm{d}x^n}[S_i(x)]\Big|_{x=0} z^i \\ &= \sum_{j=0}^{\infty} \text{Prob}[q=j] \sum_{n=0}^{j} \frac{1}{n!} \, \frac{\partial^n}{\partial x^n} \left[ \frac{1}{1-zT(x)} \right] \Big|_{x=0} \,. \end{aligned} \tag{7}$$

Working out the sum over $i$ in (7) requires that $|zT(x)| < 1$ in the neighborhood of $x = 0$, which is fulfilled for $|z| \leq 1$, since $|T(x)| < 1$ for $|x| < 1$. Considering $z$ a constant and $x$ the variable of interest, we find the partial fraction expansion

$$\frac{1}{1-zT(x)} = \sum_{p=0}^{C-1} \frac{-1}{zT'(x_p)\,(x-x_p)} \,. \tag{8}$$

By substituting (8) in (7), working out the sum over $n$, and finally again using the expansion (8)

at $x = 1$, we then obtain

$$D(z) = (z-1) \sum_{p=0}^{C-1} \frac{Q(1/x_p)}{z\,T'(x_p)\,(1-x_p)\,x_p}\,. \tag{9}$$

Thirdly, we derive a relationship between $Q(z)$ and $V(z)$. We define $f$ as the number of arrivals in slot $J$ but before P, $v_J$ as the system content at the start of slot $J$, and $t_{r,J}$ as the number of available servers from group $r$ in slot $J$. Then $q$ is given by

$$q = \max\left(0, v_J - \sum_{r=1}^{c} t_{r,J}\right) + f\,. \tag{10}$$

In order to derive $Q(z)$, we need the joint distribution of $v_J$ and $f$. This can be determined by conditioning on the value of $a_J$, the number of packet arrivals in slot $J$, as follows:

$$\text{Prob}[v_J = i, f = j] = \sum_{\ell=j+1}^{\infty} \frac{1}{\ell}\,\text{Prob}[v_J = i, a_J = \ell]\,, \tag{11}$$

since P is a random packet. Note that $\text{Prob}[v_J = i, a_J = \ell]$ corresponds to the *fraction of packets* that arrive in a slot with $\ell$ arrivals and a system content $i$ at the start of the slot. As P could be any of the $\ell$ arrivals in such a slot, $\text{Prob}[v_J = i, a_J = \ell]$ is proportional to both $b(i,\ell)$ and the number $\ell$ itself:

$$\text{Prob}[v_J = i, a_J = \ell] = \frac{\ell\, b(i,\ell)}{\lambda}\,. \tag{12}$$

The joint pgf of $v_J$ and $f$ is then obtained as

$$M(z,x) \triangleq E\left[z^{v_J} x^f\right] = \frac{B(z,1) - B(z,x)}{\lambda\,(1-x)}\,. \tag{13}$$

By means of (11), (13) and some standard $z$-transform techniques, (10) can now be transformed into

$$\begin{aligned} Q(z) = {} & T\left(\frac{1}{z}\right) M(z,z) + \frac{1}{\lambda\,(z-1)} \\ & \cdot E\left[\sum_{i=0}^{t-1} \left(1 - z^{i-t}\right) \left(\sum_{\ell=0}^{\infty} b(i,\ell)\, z^{\ell} - v(i)\right)\right]\,, \end{aligned} \tag{14}$$

where $t = \sum_{r=1}^{c} t_r$ and the expected value needs to be taken over the joint distribution of the random variables $t_1, \ldots, t_c$. On the other hand, from (1) and following similar steps, we have

$$V(z) = T\left(\frac{1}{z}\right) B(z,z) + E\left[\sum_{i=0}^{t-1} \sum_{j=0}^{\infty} b(i,j) \left(1 - z^{i-t}\right) z^j\right]. \tag{15}$$

From (13)-(15), $Q(z)$ can be expressed in terms of $V(z) = B(z,1)$. Combination of the resulting expression with (9), together with the independence of the variables $t_r$, then yields (3).

*Concluding remarks:* Relation (3) is general in the sense that the exact nature of the arrival process is not relevant. Hence, although the statistics of the system content and the delay may heavily depend on the nature of the arrival process, knowledge of this process is not needed for the transformation from system content to delay. By means of (3), not only the pgf but also several other delay characteristics, such as moments and tail probabilities, can be determined from the pgf of the system content.

Peixia Gao, Sabine Wittevrongel, Koenraad Laevens, Danny De Vleeschauwer and Herwig Bruneel (*Ghent University, Department TELIN, SMACS Research Group, Sint-Pietersnieuwstraat 41, B-9000 Gent, Belgium*)

E-mail: sw@telin.ugent.be

# References

[1] W. Whitt, A review of $L = \lambda W$ and extensions, Queueing Systems 9 (1991) 235-268.

[2] D. Bertsimas and D. Nakazato, The distributional Little's law and its applications, Operations Research 43 (1995) 298-310.

[3] Y. Xiong, H. Bruneel and B. Steyaert, Deriving delay characteristics from queue length statistics in discrete-time queues with multiple servers, Performance Evaluation 24 (1996) 189-204.

[4] K. Laevens and H. Bruneel, Delay analysis for discrete-time queueing systems with multiple randomly interrupted servers, European Journal of Operational Research 85 (1995) 161-177.

[5] P. Gao, S. Wittevrongel and H. Bruneel, Delay against system contents in discrete-time G/Geom/c queue, Electronics Letters 39 (2003) 1290-1292.