On a random area variable arising in discrete-time queues and compact directed percolation

2004 J. Phys. A: Math. Gen. 37 8421

(http://iopscience.iop.org/0305-4470/37/35/002)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND GENERAL

J. Phys. A: Math. Gen. **37** (2004) 8421–8431 PII: S0305-4470(04)80094-5

On a random area variable arising in discrete-time queues and compact directed percolation

Michael J Kearney

Advanced Technology Institute, School of Electronics and Physical Sciences, University of Surrey, Guildford, Surrey, GU2 7XH, UK

E-mail: m.j.kearney@surrey.ac.uk

Received 29 April 2004
Published 17 August 2004
Online at stacks.iop.org/JPhysA/37/8421
doi:10.1088/0305-4470/37/35/002

Abstract

A well-known discrete-time, single-server queueing system with mean arrival rate λ and mean departure rate μ is considered from the perspective of the area, A, swept out by the queue occupation process during a busy period. We determine the *exact* form of the tail of the distribution, $\Pr(A > x)$; in particular, we show that $\Pr(A > x) \sim Cx^{-1/4}\exp(-Dx^{1/2})$ for *all* $\rho \neq 1$, where $\rho \equiv \lambda/\mu$, and expressions for C and D are given. For the critical case $\rho = 1$ we show that $\Pr(A > x) \sim C'x^{-1/3}$, with C' also given. A simple mapping, used in the derivation, establishes a connection with compact directed percolation on a square lattice. As a corollary, therefore, we are also able to specify the large-area asymptotic behaviour of this model at all points in the phase diagram. This extends previous scaling results, which are only valid close to the percolation threshold.

PACS numbers: 02.50.−r, 05.50.+q, 64.60.Cn

1. Introduction

Queues are ubiquitous in the physical and social sciences, yet our understanding of certain variables associated with the queueing process is still far from complete, even for relatively simple examples (see, e.g., [1]). In this paper, we study a well-known discrete-time queueing system from the perspective of the area swept out by the queue occupation process during a busy period. It turns out that this area variable has a very interesting distribution, the tail of which can be determined exactly by exploiting a mapping onto lattice polygons. This has important applications in its own right. In addition, the same mapping establishes a connection with compact directed percolation (CDP) on a square lattice [2, 3]. Thus, as a straightforward corollary, the results obtained also determine the large-area asymptotic behaviour of this model at all points in the phase diagram. This extends previous scaling results [4, 5], which

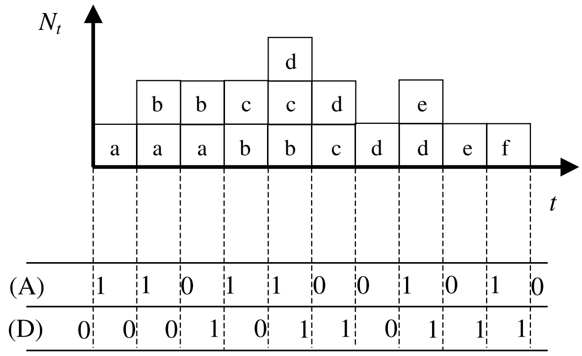

Figure 1. Queue occupancy, N_t, versus time, t, for a given queueing event. The sequences of arrivals (A) and departures (D) are indicated (early-arrival, late-departure scheme), and different 'customers' a, b, c, etc are highlighted to clarify their progress through the queue. For this event, the busy period $T = 10$ and the cumulative waiting-time $A = 17$.

are only valid for parameters tuned to be close to the percolation threshold, and also adds credence to widely held beliefs (unproven in general) about the off-critical behaviour of all percolation-type transitions [6].

The single-server, discrete-time queue in question (see figure 1) is characterized by a mean arrival rate, λ and a mean departure (or service) rate, μ. Related to these is the traffic load $\rho = \lambda/\mu$. To be precise, in any time-slot the probability of an arrival (or departure) is λ (or μ), whilst the probability of no arrival (or departure) is $\bar{\lambda} \equiv 1 - \lambda$ (or $\bar{\mu} \equiv 1 - \mu$), and arrivals and departures are independent. This is referred to as the Geo/Geo/1 queue [1], symbolizing the fact that the gaps between successive arrivals or departures are geometrically distributed. Starting from the time of the first arrival into an otherwise empty queue, the busy period, T, is the time until the queue is next empty. It is well known that $\Pr(T < \infty) = 1$ if $\rho < 1$, and the queue is stable (positive recurrent). Conversely, $\Pr(T < \infty) < 1$ when $\rho > 1$ and the queue is unstable (defective), i.e. there is a nonzero probability that T is infinite. Quantities such as the mean busy period, T_1, therefore diverge as $\rho \to 1$, and the behaviour is reminiscent of that encountered in the theory of critical phenomena. The terminology of the latter is quite natural for characterizing queueing behaviour (see also, e.g., [7]), and we identify the Geo/Geo/1 queue as being in the CDP universality class.

The area, A, swept out by the queue occupation process during a busy period is the integrated (or cumulative) waiting-time. Thus, if the busy period is the number of hours taken by the server to clear the queue, then the cumulative waiting-time is the total number of hours spent by all the 'customers' waiting for the queue to clear (the term 'customer' is used in a generic sense to describe the queueing entities, whatever they may be in reality). Both are important measures of throughput and performance, but whilst much is known about the former, very little is known about the latter. Here, we show that the tail of the area distribution, $\Pr(A > x)$, has the asymptotic behaviour $\Pr(A > x) \sim Cx^{-1/4}\exp(-Dx^{1/2})$ for all $\rho \neq 1$, and explicit expressions for C and D are given. An underlying duality means that this applies even for the case $\rho > 1$, provided that one is only concerned with the distribution of finite events. In contrast, for the critical case, $\rho = 1$, the behaviour is quite different, $\Pr(A > x) \sim C'x^{-1/3}$. Taking the continuous-time limit, we also derive the tail of the distribution for the corresponding continuous-time M/M/1 queue (M stands for Markovian). A solution for the latter, for $\rho < 1$, has been given in [8] using quite different techniques. The agreement found provides, in a satisfying way, independent confirmation of the correctness of the results, as well as their extension to $\rho \geqslant 1$ in the continuous-time case.

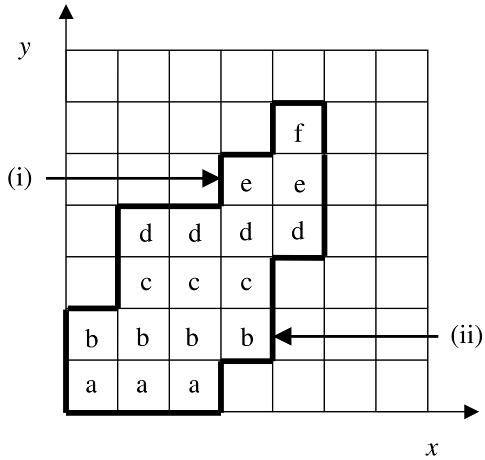

Figure 2. A staircase polygon representation of the queueing event depicted in figure 1. The upper directed walk (i) corresponds to arrivals, the lower directed walk (ii) corresponds to departures and the queue occupancy at a given time can be read-off diagonally (top–left to bottom–right). The weight of the event is $\lambda^5\bar{\lambda}^5\mu^6\bar{\mu}^4$.

It is significant that the results also offer, through a simple mapping, new insights into CDP. To set the scene, cluster or avalanche[1] models, especially those which exhibit non-equilibrium phase transitions associated with absorbing states [9, 10], are widely employed in the study of phenomena as diverse as epidemics [9], forest-fires [6], heterogeneous catalysis [9], self-organized criticality in sandpiles [11, 12], solar flares [13], traffic jams [7] and so on. Many of these models are in the directed percolation (DP) universality class [9, 10, 14] and are analytically intractable, so one has to rely extensively on simulations. On the other hand, models in the CDP universality class are often amenable to analytic treatment, which makes them of pedagogic value. The main new result concerns the asymptotic behaviour of the area probability, P_A, i.e. the cluster-size distribution, for CDP on a square lattice [2, 3]. Specifically, at *any* point away from the percolation threshold, one has that $P_A \propto A^{-\theta}\exp(-DA^{1/2})$ as $A \to \infty$, with $\theta = 3/4$. Such off-critical behaviour, with suitably generalized exponents, is conjectured to be typical of all percolation-type transitions [6]; the present case is one of the rare instances wherein this can be established rigorously. The value $\theta = 3/4$ disagrees with the value $\theta = 4/3$ given in [15] in the context of another model in the CDP universality class. This is not a trivial matter, since precisely at the percolation threshold $P_A \propto A^{-\tau}$ with $\tau = 4/3$ [5, 11]. The source of the confusion is traced to an incorrect assumption about the form of a cross-over scaling function, which is given here in full.

2. A mapping onto lattice polygons and the busy period distribution

A typical event for the discrete-time Geo/Geo/1 queue is shown in figure 1. The service protocol is first-come, first-served, and arrivals and departures within a given time-slot occur according to the early-arrival, late-departure scheme. Each time-slot is taken to be one time unit. One can represent the arrival and departure sequences as separate, directed walks on a square lattice with appropriately weighted steps (figure 2). Each arrival (or non-arrival)

[1] The distinction between 'cluster' and 'avalanche' in the present context is basically a matter of semantics; both terms are used interchangeably in this paper.

corresponds to a step $\Delta y = 1$ ($\Delta x = 1$) of the 'arrivals walk' with weight λ ($\bar{\lambda}$), apart from the first step, which has $\Delta y = 1$ and weight 1. Conversely, each departure (or non-departure) corresponds to a step $\Delta y = 1$ ($\Delta x = 1$) of the 'departures walk' with weight μ ($\bar{\mu}$), apart from the first step, which has $\Delta x = 1$ and weight 1. Both walks start at the origin. If the event is finite, the two walks will eventually meet and annihilate. The resulting structure (figure 2) is known as a staircase polygon [16–18].

For a given queueing event, we define the activity of its associated staircase polygon to be $(\bar{\lambda}\bar{\mu}x')^{i/2}(\lambda\mu y')^{j/2}z^A$, where i, j are the total number of perimeter steps in the x and y directions, respectively, and x', y' are dummy variables which keep track of the polygon width ($i/2$), height ($j/2$) and perimeter length ($i + j$). The busy period duration, T, is given by $T = (i + j - 2)/2$, whilst the cumulative waiting-time, A, is just the area of the polygon. By construction, the probabilistic weight of an event is determined by setting $x' = y' = 1$ and $z = 1$ in the corresponding activity, and multiplying by $(\lambda\bar{\mu})^{-1}$. The latter factor arises because the initial arrival is assumed to occur with probability 1. It follows that all the quantities of interest may be obtained from the area–perimeter generating function for staircase polygons [4]. For the present purpose, one can set $y' = x'$ without loss of generality.

It is natural to define a (probability) generating function, $\tilde{G}(x', z)$, which is related to the joint probability, $P(T, A)$, as follows:

$$\tilde{G}(x', z) \equiv \lambda\bar{\mu} \sum_{T,A} P(T, A) x'^{T+1} z^A. \tag{1}$$

In (1), the sum is taken over all possible finite events. Based on the connection with staircase polygons, one can show that $\tilde{G}(x', z)$ obeys a nonlinear functional equation of quadratic type [4, 17],

$$\tilde{G}(x', z) = \alpha x'^2 z + \beta x' z \tilde{G}(x', z) + \tilde{G}(x', z)\tilde{G}(x'z, z) \tag{2}$$

where $\alpha = \lambda\bar{\lambda}\mu\bar{\mu}$ and $\beta = \lambda\mu + \bar{\lambda}\bar{\mu}$. This functional approach is both elegant and powerful; for example, many results in queueing theory relating to the busy period, T, can be deduced from (2) very quickly by setting $z = 1$. We illustrate some of these below, but with the main objective of highlighting certain mathematical structures which will be useful in what follows. The asymptotic behaviour of the area distribution, on the other hand, is controlled by $\tilde{G}(1, z)$ in the limit $z \to 1^-$. This limit is much more subtle and its discussion is deferred until the next section.

From (1) and (2) we have the following result for the busy period probability, P_T, and its associated generating function, $\tilde{G}(x', 1)$,

$$\tilde{G}(x', 1) \equiv \lambda\bar{\mu} \sum_T P_T x'^{T+1} = \frac{1 - \beta x' - \sqrt{(1 - \beta x')^2 - 4\alpha x'^2}}{2}. \tag{3}$$

Since the sum is over finite events, $\Pr(T < \infty) = (\lambda\bar{\mu})^{-1}\tilde{G}(1, 1)$. There are two distinct regimes to discuss. When $\Delta \equiv \lambda - \mu < 0$, it follows from (3) that queueing events are finite with probability 1. Conversely, infinite events can occur when $\Delta > 0$ with probability $P_\infty \equiv 1 - \Pr(T < \infty) = (\lambda\bar{\mu})^{-1}\Delta$ [3]. The moments, T_k, can be obtained in the usual manner by differentiating (3) repeatedly with respect to x'. For the mean busy period one finds that $T_1 = (\mu - \lambda)^{-1}$ for $\Delta < 0$. For $\Delta > 0$, the relevant quantity to consider is the mean busy period *given* that the events are finite; in other words, one needs to condition P_T by dividing by $\Pr(T < \infty) = 1 - P_\infty$. The result is that $T_1 = (\lambda - \mu)^{-1}$, and thus $T_1 = |\Delta|^{-1}$ for all $\lambda \neq \mu$. This illustrates an important general principle; statistical quantities below ($\Delta < 0$) and above ($\Delta > 0$) the critical point are related by the duality transformation $\lambda \leftrightarrow \mu$. This is discussed further in section 4. Note that α, β and $|\Delta|$ are all invariant under this transformation. The introduction of a further variable, η, which takes the value $\eta = \lambda\bar{\mu}$ when $\Delta < 0$ and $\eta = \bar{\lambda}\mu$ when $\Delta > 0$, i.e. $\eta = \min\{\lambda\bar{\mu}, \bar{\lambda}\mu\}$, means that all results can be written in a form which is valid either side of the critical point. Regarding the probability, P_T, it is possible to invert (3) explicitly,

$$P_T = \frac{\beta^{T+1}}{2\eta(T+1)\xi^{T+1}}[\xi L_T(\xi) - L_{T-1}(\xi)]; \qquad \xi \equiv \frac{\beta}{(\beta^2 - 4\alpha)^{1/2}} \tag{4}$$

where $L_T(\xi)$ is the Tth Legendre polynomial. This expression is exact for all $T \geqslant 1$. Of particular interest, however, is the asymptotic form of P_T as $T \to \infty$, and this can be evaluated using standard results from the theory of Legendre functions

$$P_T \sim \frac{\alpha^{1/4}(\beta + 2\sqrt{\alpha})^{1/2}}{2\sqrt{\pi}\eta} T^{-3/2} \exp\left\{-T \log\left(\frac{1}{\beta + 2\sqrt{\alpha}}\right)\right\}. \tag{5}$$

Here, and throughout the paper, the statement $X \sim Y$ is made in the strict sense that $X/Y \to 1$ in a specified limit. Several things about (5) are worth noting. First, it is valid for all λ, μ. Second, $\beta + 2\sqrt{\alpha} \leqslant 1$, and the equality is only attained when $\lambda = \mu$. Thus, one has pure power-law behaviour at the critical point. Third, the tail of the survival probability, $\Pr(T > t)$, is of the form $\Pr(T > t) \sim t^{-1/2} g(\zeta t^{1/2})$, where $\zeta = [-\log(\beta + 2\sqrt{\alpha})]^{1/2}$. The function $g(s)$ has the following limits:

$$g(s) \sim \frac{\alpha^{1/4}(\beta + 2\sqrt{\alpha})^{1/2}}{2\sqrt{\pi}\eta} s^{-2} \exp\{-s^2\}; \qquad s \to \infty \tag{6}$$

$$g(0) = \frac{1}{\sqrt{\pi \mu\bar{\mu}}}. \tag{7}$$

In the terminology of critical phenomena this is typical, cross-over scaling behaviour, except that there is *no* requirement here that $|\Delta|$ is small. However, it is only in limit $|\Delta| \to 0$, when $\eta \to \mu\bar{\mu}$ and $\zeta \to |\Delta|/2(\mu\bar{\mu})^{1/2}$, that one can use (5) to calculate the asymptotic behaviour of the moments,

$$T_k \sim \frac{\Gamma\left(k - \frac{1}{2}\right)4^{k-1}}{\sqrt{\pi}} \frac{(\mu\bar{\mu})^{k-1}}{|\lambda - \mu|^{2k-1}}. \tag{8}$$

This result, which can also be derived by directly manipulating (3), holds for all $k \geqslant 1$. Conversely, given (8), one can uniquely determine (5) in the limit $|\Delta| \to 0$.

3. Asymptotic behaviour of the area variable

The mean area, A_1, can be obtained by differentiating (2) with respect to z. The result is $A_1 = (\alpha/\eta)|\lambda - \mu|^{-2}$. The principal objective of the present work, however, is to establish the asymptotic behaviour of the tail of the area probability distribution, $\Pr(A > x)$, which is determined by the singular part of $\tilde{G}(1, z)$ as $z \to 1^-$. The latter has been obtained from a formal q-series solution of (2) after a long and difficult analysis [18]. In the present context, the key result can be written as

$$\tilde{G}(1, z) \sim \lambda\bar{\mu} - \frac{\Delta}{2} + \frac{|\Delta|}{2\sqrt{\kappa}}(1 - z)^{1/3} \frac{\mathrm{Ai}'(\kappa(1 - z)^{-2/3})}{\mathrm{Ai}(\kappa(1 - z)^{-2/3})} \tag{9}$$

where $\mathrm{Ai}(s)$ is the Airy function and $\mathrm{Ai}'(s)$ is its derivative, and

$$\frac{4}{3}\kappa^{3/2} = \log(\bar{\lambda}\bar{\mu}) \log\left(\frac{\lambda}{\mu}\right) + 2\mathrm{Li}_2(\lambda) - 2\mathrm{Li}_2(\mu) \tag{10}$$

where $\mathrm{Li}_2(s)$ is the dilogarithm function and it is understood that $\kappa \geqslant 0$ is real. Note that $\kappa = 0$ when $\lambda = \mu$. With reference to (1), the asymptotic behaviour of the area probability, P_A, can be evaluated by inverting (9) using contour integration techniques. The basic method for problems of this type is covered in [16, 19], with the result that for $A \to \infty$,

$$P_A \sim \frac{|\Delta|}{2\eta\sqrt{\kappa}} A^{-4/3} f(\sqrt{\kappa} A^{1/3}) \tag{11}$$

$$f(s) = -\sum_{k=0}^{\infty} \frac{C_k}{\Gamma\left(\frac{2k-1}{3}\right)} s^{2k}; \qquad f(s) > 0. \tag{12}$$

The C_k in (12) are the Taylor series coefficients of $F(s) \equiv -\mathrm{Ai}'(s)/\mathrm{Ai}(s)$ when expanded about the origin. Using the fact that $F(s)$ obeys a Riccati equation, $F(s)^2 - F'(s) - s = 0$, we have

$$C_{k+1} = \frac{1}{k+1} \sum_{i=0}^{k} C_i C_{k-i}; \qquad k \geqslant 2 \tag{13}$$

with $C_0 = 3^{1/3}\Gamma\left(\frac{2}{3}\right)\big/\Gamma\left(\frac{1}{3}\right)$, $C_1 = C_0^2$ and $C_2 = C_0 C_1 - \frac{1}{2}$. It is important to stress that the function $f(s)$ is entire and that (11) holds for all λ, μ.

The above results are relatively straightforward extensions of what is already known. What has not been identified to date, however, is the precise asymptotic behaviour of $f(s)$ as $s \to \infty$. This is given by

$$f(s) \sim \frac{1}{\sqrt{\pi}} \left(\frac{4}{3}\right)^{1/4} s^{7/4} \exp\left(-\left(\frac{16}{3}\right)^{1/2} s^{3/2}\right). \tag{14}$$

We shall discuss the derivation of this result shortly. Accepting it for now, it follows that for $\lambda \neq \mu$ (i.e. $\kappa \neq 0$) and $A \to \infty$,

$$P_A \sim \frac{1}{\sqrt{2\pi}} \left(\frac{1}{3}\right)^{1/4} \frac{|\Delta|\kappa^{3/8}}{\eta} A^{-3/4} \exp\left(-\left(\frac{16}{3}\right)^{1/2} \kappa^{3/4} A^{1/2}\right). \tag{15}$$

With regard to the tail of the area probability distribution, one therefore has

$$\Pr(A > x) \sim \frac{1}{2\sqrt{\pi}} \left(\frac{3}{4}\right)^{1/4} \frac{|\Delta|}{\eta\kappa^{3/8}} x^{-1/4} \exp\left(-\left(\frac{16}{3}\right)^{1/2} \kappa^{3/4} x^{1/2}\right). \tag{16}$$

This is of the form $\Pr(A > x) \sim C x^{-1/4} \exp(-D x^{1/2})$ for *all* $\lambda \neq \mu$. The distribution has a heavy (Weibull-like) tail, and it is somewhat surprising to encounter such rich behaviour in so simple a system. For $\lambda = \mu$ one can take the limit $|\Delta| \to 0$ in (11), noting that $\kappa \to |\Delta|^2/4(\mu\bar{\mu})^{4/3}$, to give

$$\Pr(A > x) \sim \frac{3^{1/3}}{\Gamma\left(\frac{1}{3}\right)(\mu\bar{\mu})^{1/3}} x^{-1/3}. \tag{17}$$

These results are of direct relevance to a variety of applications encountered within the traditional domain of queueing theory (ranging from communication networks to inventory processes) [1]. The key message is that 'area' distributions appear to be naturally long tailed, with important consequences for the likelihood of 'extreme' events occurring.

It remains to establish (14). The direct approach, using, e.g., steepest descent techniques, is technically challenging. Fortunately, an alternative method is available. This involves deriving P_A for $|\Delta|$ small but nonzero, using the moments as a guide, and deducing (14), and hence (15) and (16), retrospectively. The argument runs as follows. In the limit $s \to \infty$, the function $F(s) \equiv -\mathrm{Ai}'(s)/\mathrm{Ai}(s)$ has the following asymptotic expansion [20]:

$$F(s) \sim s^{1/2} \sum_{k=0}^{\infty} a_k s^{-3k/2} = s^{1/2} + \frac{1}{4} s^{-1} - \frac{5}{32} s^{-5/2} + \cdots \tag{18}$$

where the coefficients a_k obey a quadratic recurrence relation

$$a_k = -\frac{1}{2} \sum_{i=1}^{k-1} a_i a_{k-i} - \left(\frac{3k}{4} - 1\right) a_{k-1}; \qquad k \geqslant 2 \tag{19}$$

with $a_0 = 1$ and $a_1 = 1/4$. A closed-form expression for a_k is not available. However, by using results that relate $F(s)$ to certain hypergeometric functions [20], one can derive in the limit $k \to \infty$,

$$a_k \sim (-1)^{k+1} \sqrt{2} \left(\frac{3}{256}\right)^k \frac{(4k)!}{(k)!(2k)!}. \tag{20}$$

From (9) and (18), one establishes that $\tilde{G}(1, 1-\varepsilon)$ has an asymptotic expansion in integer powers of $\varepsilon \equiv 1 - z$. Further, in the limit $|\Delta| \to 0$ the divergence of the area moments, A_k, can be evaluated asymptotically for all $k \geqslant 1$ as

$$A_k \sim \frac{(-1)^k}{\mu\bar{\mu}} \frac{\partial^k \tilde{G}(1, 1-\varepsilon)}{\partial \varepsilon^k}\bigg|_{\varepsilon=0} \sim (-1)^{k+1} a_k 2^{3k-1} k! \frac{(\mu\bar{\mu})^{2k-1}}{|\lambda - \mu|^{3k-1}}. \tag{21}$$

From this and (20), it follows that in the twin limits $|\Delta| \to 0$ and $k \to \infty$,

$$A_k \sim \frac{1}{\sqrt{2\pi}} \left(\frac{3}{2}\right)^k \Gamma\left(2k + \frac{1}{2}\right) \frac{(\mu\bar{\mu})^{2k-1}}{|\lambda - \mu|^{3k-1}}. \tag{22}$$

This result, interesting in its own right, is sufficient to determine the asymptotic behaviour of P_A for $|\Delta| \approx 0$. We note that *any* distribution whose moment sequence has the property (22) is *uniquely* determined by its moment sequence. The relevant test is known as the Carleman criterion [21, 22], i.e. $\Sigma_{k\geqslant 1} A_k^{-1/2k} = \infty$, which is seen to be satisfied since, from (22), $A_k^{-1/2k} = O(k^{-1})$ as $k \to \infty$. An extension of the basic reasoning leads to the stronger statement that all such distributions are asymptotically equivalent as $A \to \infty$; in other words, the asymptotic behaviour of P_A is uniquely determined by (22). Consider, now, the ansatz $P_A \sim \hat{C} A^{-\theta} \exp(-\hat{D} A^{\xi})$. By evaluating the moments in relation to (22) one finds that such a solution exists, namely,

$$P_A \sim \frac{1}{2\sqrt{\pi}} \left(\frac{1}{6}\right)^{1/4} \frac{|\lambda - \mu|^{7/4}}{(\mu\bar{\mu})^{3/2}} A^{-3/4} \exp\left\{-\left(\frac{2}{3}\right)^{1/2} \frac{|\lambda - \mu|^{3/2}}{\mu\bar{\mu}} A^{1/2}\right\}. \tag{23}$$

By uniqueness, this is the required asymptotic solution for fixed $|\Delta| \approx 0$ and $A \to \infty$. Comparison with (11), noting that $\kappa \approx |\Delta|^2/4(\mu\bar{\mu})^{4/3}$ for $|\Delta| \approx 0$, then shows that, asymptotically, $f(s)$ is given by (14) as $s \to \infty$. But, since $f(s)$ itself is independent of Δ, this result must be independent of the assumption that $|\Delta| \approx 0$. One can then use (14) to establish (15) and (16) for all values of κ.

Confirmation of (16) by another route comes from considering the equivalent problem for the continuous-time version of the Geo/Geo/1 queue, which is referred to as the M/M/1 queue [1]. This has recently been solved for $\rho < 1$ using completely different techniques, i.e. by analysing a continued fraction representation of the Laplace transform of the distribution [8]. By taking the continuous-time limit here we can also derive such results. To do so one simply needs to introduce the time-step, Λ, into the above analysis through the substitutions

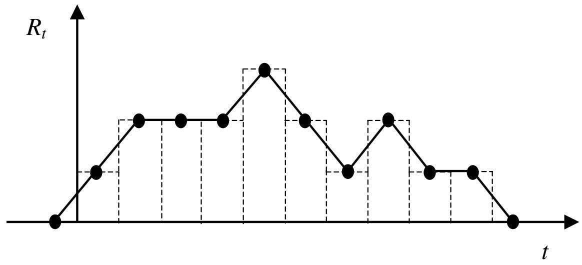

Figure 3. An alternative representation of figure 1, establishing a connection with a random walk on a directed triangular lattice (solid line).

$\lambda \to \hat{\lambda}\Lambda$, $\mu \to \hat{\mu}\Lambda$ and $A \to \hat{A}\Lambda^{-1}$, and let $\Lambda \to 0$. When the traffic load $\rho = \hat{\lambda}/\hat{\mu} < 1$, one has

$$\Pr(\hat{A} > x) \sim \frac{1-\rho}{\rho\sqrt{2\pi\gamma}} x^{-1/4} \exp\{-\gamma x^{1/2}\} \tag{24}$$

where $\gamma = 2\hat{\mu}^{1/2}\sqrt{-2(1-\rho) - (1+\rho)\log\rho}$. This is the *exact* result obtained in [8], which establishes the correctness of (16), and hence (14). The analysis in [8] does not, however, provide information about the null recurrent case, $\rho = 1$, or the defective case, $\rho > 1$. The latter (for finite events) is given by (24) under the transformation $\hat{\lambda} \leftrightarrow \hat{\mu}$ and $\rho \leftrightarrow \rho^{-1}$, whilst the former is given by taking the continuous-time limit of (17). Thus, as well as verifying existing results, the analysis for the Geo/Geo/1 queue adds to what is known about the area distribution in the M/M/1 queue.

We end this section by remarking that the queue occupation process for the Geo/Geo/1 queue is equivalent to a random walk on a directed triangular lattice. Consider figure 3, which is a redrawing of figure 1. One is basically adding successive columns whose height, in relation to the preceding column, can increase or decrease by one unit, or stay the same. Identifying the top of each column with successive steps then defines a random walk, R_t, with transition probabilities $P_+ = \lambda\bar{\mu}$, $P_- = \bar{\lambda}\mu$ and $P_0 = \lambda\mu + \bar{\lambda}\bar{\mu}$. The results given above then relate to the time of first return, $T' = T + 1$, and the area swept out, $A' = A$, of such a process, noting that $\alpha = P_+P_-$ and $\beta = P_0$. This points to other applications. For example, if R_t denotes the random evolution of the power in some burst (eruption) process, then A' is simply the total energy released in the burst. Or, if R_t denotes the random excess strain or overload (beyond the critical fatigue threshold) in a mechanical system, then A' is a measure of the cumulative fatigue damage incurred, which relates to the lifetime of the system. Of course, this model is only an abstraction of such real world problems, but it serves to highlight features one might expect to observe in reality regarding the tail of the relevant distributions.

4. The link to compact directed percolation

CDP, which is closely related to the voter model and to the Glauber–Ising model at zero temperature [9], has been widely studied. It was originally proposed by Domany and Kinzel in the context of a cellular automaton [2]. The determination of certain critical exponents and hyper-scaling relations followed [3, 23], and further extensions have included modified models to reflect the presence of impenetrable walls and other barriers [24–30]. In this context, we also note the avalanche model of [15] which, although based around the area swept out by Dyck paths, is in the CDP universality class. Recently, the formal connection with staircase polygon models has led to a deeper understanding of the scaling behaviour and of certain

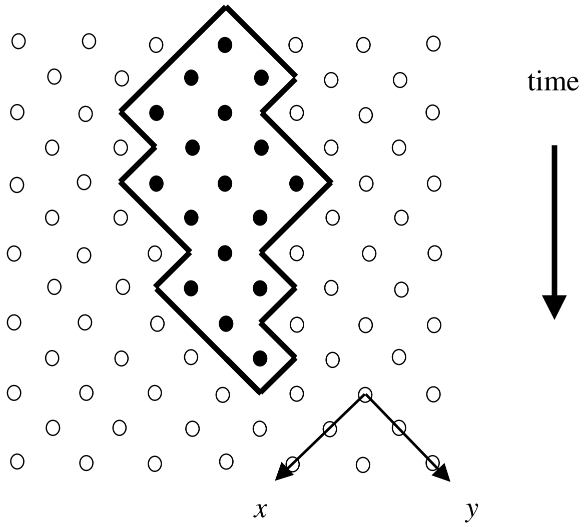

Figure 4. A cluster or avalanche event (full circles) of weight $p_x^4 q_x^6 p_y^5 q_y^5$. The duration of the event $T = 10$ and the size or area $A = 17$.

scaling functions [4, 5]. However, these results, obtained by analysing (2) using dominant balance techniques [31], are only valid near the percolation threshold.

The model is defined as follows [2–5]. At time $t = 0$, a single site in a directed square lattice becomes active with probability 1. At integer time $t > 0$, sites are updated depending upon the states of their nearest neighbour preceding sites (figure 4). If one or other (but not both) of the preceding sites is active, then the site becomes active with probability p_x or p_y, depending upon which direction is invoked. However, if both preceding sites are active, then the site becomes active with probability 1. This rule ensures that the clusters are compact. With reference to figure 4, for finite events the cluster size is the area, A, of the (staircase) polygon defined by the closed walk (on the dual lattice) which bounds the active sites as tightly as possible. The connection between this process and the queueing process depicted in figure 2 is self-evident. In fact, all the results obtained above for the Geo/Geo/1 queue map over through the simple replacement $\lambda = p_y$ and $\bar{\mu} = p_x$. The percolation threshold is then given by $\Delta \equiv p_x + p_y - 1 = 0$, i.e. $p_x + p_y = 1$ [3]. Behaviour below ($\Delta < 0$) and above ($\Delta > 0$) this threshold is related by the transformation $p_x \leftrightarrow q_y$ and $p_y \leftrightarrow q_x$. This has been clearly explained in [3] in terms of a fundamental duality between active and inactive sites, the key point being that there are actually *two* absorbing states: all sites active and all sites inactive (see also [9]).

It suffices to summarize the main new results that follow from the identified mapping. Regarding the area probability, P_A, this is given by (15) for all $|\Delta| \neq 0$. Thus, one has $P_A \propto A^{-\theta} \exp(-DA^{1/2})$ as $A \to \infty$, with $\theta = 3/4$. It is interesting that this scaling form is also seen in relation to the partition function for models of polymer collapse [32], with which there appears to be a deep connection. Near to the percolation threshold, i.e. $|\Delta| \approx 0$, one has the striking result

$$P_A \sim \frac{1}{2\sqrt{\pi}} \left(\frac{1}{6}\right)^{1/4} \frac{|p_x + p_y - 1|^{7/4}}{(p_x p_y)^{3/2}} A^{-3/4} \exp\left\{-\left(\frac{2}{3}\right)^{1/2} \frac{|p_x + p_y - 1|^{3/2}}{p_x p_y} A^{1/2}\right\}. \qquad (25)$$

It should be noted that a value of $\theta = 4/3$ was proposed in [15], based on incomplete knowledge about the behaviour of $f(s)$ as $s \to \infty$. In particular, the prefactor $s^{7/4}$ in (14) was missing. However, as correctly identified in [15], and before that in [11], at the percolation threshold itself one expects $P_A \propto A^{-4/3}$, and indeed from (11) it follows that

$$P_A \sim \frac{A^{-4/3}}{3^{2/3}\Gamma\left(\frac{1}{3}\right)(p_x p_y)^{1/3}}; \qquad |\Delta| = 0. \tag{26}$$

With these results, the asymptotic behaviour of the cluster-size distribution is determined at all points in the phase diagram. There is a wider significance to this. For standard percolation, controlled, say, by a single lattice parameter, p, whether directed or undirected, it has been proved that as $A \to \infty$, $\log P_A \propto -A$ for $p \ll p_c$ and $\log P_A \propto -A^{1-1/d}$ for $p \gg p_c$, where p_c is the percolation threshold and d is the dimensionality [33–35]. This has led to the conjecture that the off-critical behaviour in percolation is of the form $P_A \propto A^{-\theta_\pm} \exp(-D_\pm A^{\xi_\pm})$ for *all* $p \neq p_c$ as $A \to \infty$. The present problem provides a rare instance wherein this can be shown to be true, albeit with $\xi_- = 1/2$ (not $\xi_- = 1$) below the percolation threshold, which is a reflection on the compact nature of the model. The fact that $\xi_+ = 1/2$ suggests that the (rare) clusters that contribute in standard two-dimensional percolation for $p > p_c$ are, essentially, 'compact' in nature [6].

Viewing the clusters as avalanches, their duration can be characterized using the results given in section 2 after the replacement $\lambda = p_y$ and $\bar{\mu} = p_x$. Again, these are valid everywhere, not just near to the percolation threshold, which extends what has been given previously in the literature.

5. Conclusions

In this paper, the area swept out by the queue occupation process during a busy period has been considered for the discrete-time Geo/Geo/1 queue and its continuous-time counterpart, the M/M/1 queue. In particular, the tail of the area distribution has been determined exactly. The analysis is based on the study of a nonlinear functional equation, and it is clear that this approach provides new and powerful insights into the behaviour of queueing systems in general. Of particular significance is the fact that no assumption has been made about the relative values of the arrival and departure rates. From an applications perspective, it is relevant that the distribution has a heavy tail.

A simple mapping enables the derived results to be applied immediately to compact directed percolation on a square lattice. For the cluster-size distribution, P_A, the most important new finding has been to show that $P_A \propto A^{-3/4} \exp(-DA^{1/2})$ as $A \to \infty$ at *all* points in the phase diagram, except precisely at the percolation threshold. This extends previous scaling results, which are only valid near to the percolation threshold, and provides one of the most comprehensive analyses to date for any percolation-type problem.

References

[1] Asmussen S 2003 *Applied Probability and Queues* 2nd edn (New York: Springer)
[2] Domany E and Kinzel W 1984 *Phys. Rev. Lett.* **53** 311
[3] Essam J W 1989 *J. Phys. A: Math. Gen.* **22** 4927
[4] Kearney M J 2002 *J. Phys. A: Math. Gen.* **35** L731
[5] Kearney M J and Martin R J 2003 *J. Phys. A: Math. Gen.* **36** 6629
[6] Stauffer D and Aharony A 1994 *Introduction to Percolation Theory* 2nd edn (London: Taylor and Francis)
[7] Ha M, Timonen J and den Nijs M 2003 *Phys. Rev.* E **68** 056122
[8] Guillemin F and Pinchon D 1998 *Queueing Syst.* **29** 383

[9] Hinrichsen H 2000 *Adv. Phys.* **49** 815
[10] Odor G 2002 *Preprint* cond-mat/0205644 (2004 *Rev. Mod. Phys.* at press)
[11] Dhar D and Ramaswamy R 1989 *Phys. Rev. Lett.* **63** 1659
[12] Kloster M, Maslov S and Tang C 2001 *Phys. Rev.* E **63** 026111
[13] Charbonneau P, McIntosh S W, Liu H-L and Bogdan T J 2001 *Solar Phys.* **203** 321
[14] Mohanty P K and Dhar D 2002 *Phys. Rev. Lett.* **89** 104303
[15] Jonsson T and Wheater J F 1998 *J. Stat. Phys.* **92** 713
[16] van Rensburg E J J 2000 *The Statistical Mechanics of Interacting Walks, Polygons, Animals and Vesicles* (Oxford: Oxford University Press)
[17] Prellberg T and Brak R 1995 *J. Stat. Phys.* **78** 701
[18] Prellberg T 1995 *J. Phys. A: Math. Gen.* **28** 1289
[19] Brak R and Owczarek A L 1995 *J. Phys. A: Math. Gen.* **28** 4709
[20] Flajolet P and Louchard G 2001 *Algorithmica* **31** 361
[21] Kendall M G, Stuart A and Ord J K 1994 *Kendall's Advanced Theory of Statistics* vol 1: *Distribution Theory* 6th edn (London: Arnold)
[22] Pakes A G, Hung W-L and Wu J-W 2001 *Aust. N.Z. J. Stat.* **43** 101
[23] Dickman R and Tretyakov A Y 1995 *Phys. Rev.* E **52** 3218
[24] Lin J -C 1992 *Phys. Rev.* A **45** R3394
[25] Essam J W and TanlaKishani D 1994 *J. Phys. A: Math. Gen.* **27** 3543
[26] Essam J W and Guttmann A J 1995 *J. Phys. A: Math. Gen.* **28** 3591
[27] Brak R and Essam J W 1999 *J. Phys. A: Math. Gen.* **32** 335
[28] Mendes J F F, Dickman R and Hermann H 1996 *Phys. Rev.* E **54** R3071
[29] Atman A P F, Dickman R and Moreira J G 2002 *Phys. Rev.* E **66** 016113
[30] Kearney M J 2002 *J. Phys. A: Math. Gen.* **35** L421
[31] Richard C 2002 *J. Stat. Phys.* **108** 459
[32] Owczarek A L 1993 *J. Phys. A: Math. Gen.* **26** L647
[33] Kunz H and Souillard B 1978 *Phys. Rev. Lett.* **40** 133
[34] Harris A B and Lubensky T C 1981 *Phys. Rev.* B **24** 2656
[35] Lubensky T C and McKane A J 1981 *J. Phys. A: Math. Gen.* **14** L157