Processor-Ring Communication: A Tight Asymptotic Bound on Packet Waiting Times

E. G. Coffman, Jr.,[1] *Nabil Kahale*[2] and *F. T. Leighton*[3]

[1]AT&T Bell Laboratories, Murray Hill, NJ 07974
[2]DIMACS, Rutgers University, New Brunswick, NJ 08855
[3]Dept. of Math. and Laboratory for Computer Science, MIT, Cambridge, MA 02139

July 13, 1999

ABSTRACT

We consider N processors communicating unidirectionally over a closed transmission channel, or ring. Each message is assembled into a fixed-length packet. Packets to be sent are generated at random times by the processors, and the transit times spent by packets on the ring are also random. Packets being forwarded, i.e., packets already on the ring, have priority over waiting packets. The objective of this paper is to analyze packet waiting times under a greedy policy, within a discrete Markov model that retains the over-all structure of a practical system, but is simple enough so that explicit results can be proved. Independent, identical Bernoulli processes model message generation at the processors, and i.i.d. geometric random variables model the transit times. Our emphasis is on asymptotic behavior for large ring sizes, N, when the respective rate parameters have the scaling λ/N and μ/N. Our main result shows that, if the traffic intensity is fixed at $\rho = \lambda/\mu < 1$, then as $N \to \infty$ the expected time a message waits to be put on the ring is bounded by a constant. This result verifies that the expected waiting time under the greedy policy is within a constant factor of that under an optimal policy.

Processor-Ring Communication: A Tight Asymptotic Bound on Packet Waiting Times

E. G. Coffman, Jr.,[1] *Nabil Kahale*[2] and *F. T. Leighton*[3]

[1]AT&T Bell Laboratories, Murray Hill, NJ 07974
[2]DIMACS, Rutgers University, New Brunswick, NJ 08855
[3]Dept. of Math. and Laboratory for Computer Science, MIT, Cambridge, MA 02139

1. Introduction

Communication among M processors takes place counterclockwise along a slotted circular transmission channel, or *ring*. A processor generates messages, receives messages, and forwards messages between other processors. Each message is a *packet* of fixed duration. One time unit is required for a packet to be sent or forwarded from one processor to its counterclockwise neighbor. Packets are generated randomly at the processors according to i.i.d. arrival processes. The integer times spent by packets on the ring, packet *transit times*, are i.i.d. random variables. Packets being forwarded on the ring have priority: while a processor has a packet to be forwarded, it can not place one of its own waiting packets on the ring. A packet waiting for transmission is held in a queue at the processor where it was generated.

The details defining a practical implementation of a processor ring are many and varied. Indeed, the applications and analysis of communication rings form a rather large and growing literature; see van Arem and van Doorn (1990), Barroso and Dubois (1993), and Georgiadis, Szpankowski, and Tassiulas (1993) for brief surveys and many references. As a concession to mathematical tractability, we adopt here the simple discrete Markov model in Fig. 1, where the ring is partitioned into *cells*, each capable of holding a single packet. The cells rotate counterclockwise past the processors in discrete steps, 1 step per unit of time. Packets are generated at each of the N processors by a Bernoulli process at rate λ/N $0 < \lambda < N$, per time unit (step); the total arrival rate is then λ. The packet transit times are geometrically distributed with rate parameter μ/N, $N > \mu > \lambda$. Thus, at any given step, a packet on the ring departs with probability μ/N and stays for at least one more step with probability $1-\mu/N$, independent of how long the packet has already been on the ring. We will explain shortly the reason for the scaling of arrival and transit-time parameters by the ring size.

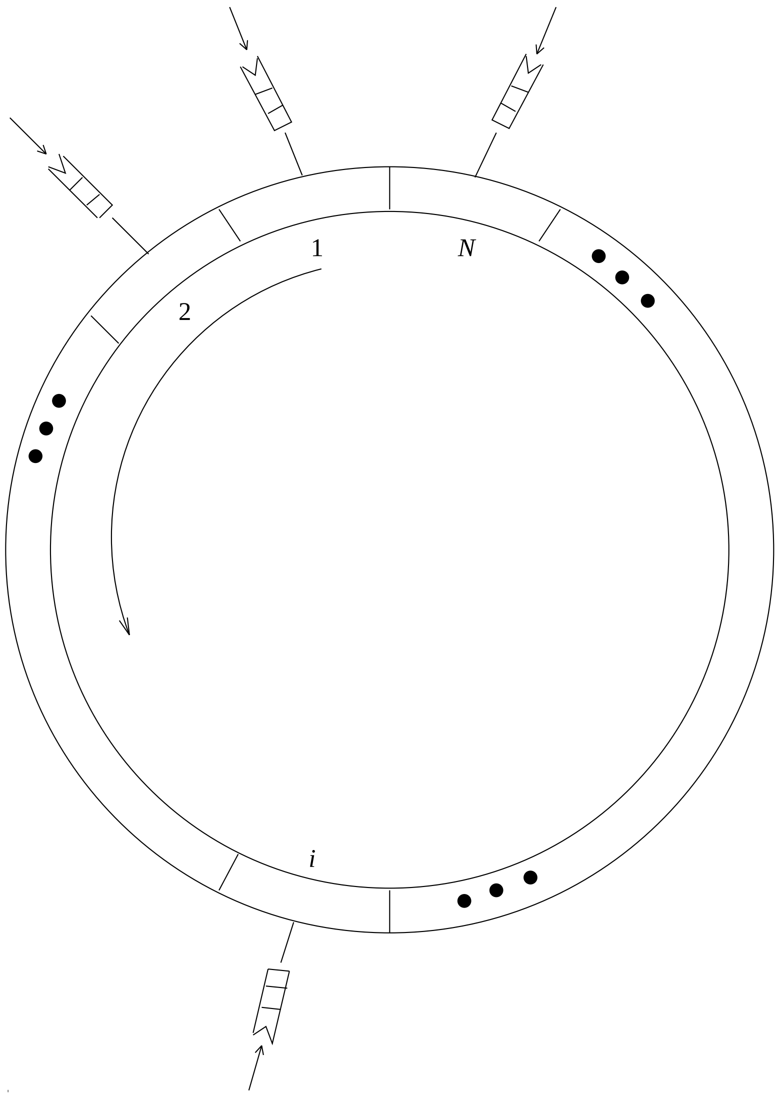

Figure 1: The rotating ring model.

In each step, the ring system undergoes a transition according to the following sequence:

(i) The ring rotates one position while processor queues accept new arrivals, if any (at most 1 per queue in each step).

(ii) Packets on the ring that are opposite their destinations are delivered, i.e., removed from their cells.

(iii) Each processor with a nonempty queue opposite an empty cell then puts a waiting packet into this cell.

This gives the *nonblocking* model; reversing (ii) and (iii) would give the *blocking* model: a departing packet can not be replaced in the same time step by a waiting packet. As we shall see, our asymptotic results apply to both models. The above sequence gives the *greedy* cell admission policy, placing waiting packets on the ring as soon as empty cells are available.

As discussed in Coffman et al. (1993), the greedy policy has the undesirable effect of occasionally "freezing out" certain processor queues for long periods of time; long trains of occupied cells pass by such processors denying them access to the ring. The results of this paper will show that, for large rings within our probability model, the greedy rule is remarkably efficient, and that in fact the above behavior is quite rare.

Our specific objective is to analyze packet waiting times under the greedy policy. (Hereafter, unless noted otherwise, waiting times always refer to times spent waiting in processor queues.) An exact analysis of the ring system appears to be quite difficult. A Markov-chain approach suggests the difficulties to be encountered; the state of a Markov chain must include the number in each queue and the state, occupied or empty, of each cell of the ring. Thus, we turn to asymptotic estimates for large ring sizes, N, with λ and μ fixed. This is why we introduced the scalings λ/N and μ/N; as we allow N to increase, the traffic intensity will remain fixed at $\rho = \lambda/\mu$, the usual product of arrival rate and average service (transit) time. To prepare for our main theorem, we need a little more notation. Let $Q_i^N(t)$ denote the number in the i^{th} processor queue at integer time $t \geq 0$ in a ring of N cells. Let Q^N have the stationary distribution common to all queue lengths $Q_i^N(t)$, assuming that it exists. Let W^N be the waiting time of a packet in the stationary regime.

Theorem 1.1. *Fix λ and μ with $\lambda < \mu$. Then the stationary distribution of $Q_i^N(t)$ exists and has an expectation*

$$E[Q^N] = \Theta(1/N) \ .$$

Thus, by Little's theorem,

$$E[W^N] = \Theta(1) \ .$$

The fact that the ring process is ergodic when $\lambda < \mu$ and hence $\rho < 1$ has already been proved by Coffman, et al. (1993). Also, there is no need to spend time on the proof of the lower bounds, for these are easy to see, as follows. Consider the entire ring as an N-server system with a total arrival rate λ and maximum departure rate μ. Then by Little's theorem, the arrival rate λ times the average time spent on the ring, i.e., N/μ, must be equal to the expected number of packets on the ring in the stationary regime, i.e., ρN. It is easily seen that if a fraction $\rho > 0$ of the ring is occupied on average, then $E[W^N] = \Omega(1)$ and hence $E[Q^N] = \Omega(1/N)$.

This paper is a sequel to the work of Coffman et al. (1993) who proved the weaker theorem $E[Q^N] = o(1)$, $N \to \infty$, as their main result. They also presented results of an experimental study, which led to interesting conjectures and open problems, the problem solved here being one of them. Theorem 1.1 on the convergence rate uses the same combinatorial set-up, which is presented in the next section, but the probabilistic analysis here is far more intricate. The law of large numbers was the basic tool in Coffman et al. (1993). Here, however, we will need more powerful asymptotic bounds (e.g., those of Chernoff type) on the tail probabilities for sums of independent random variables and the excursions of Lindley processes; these appear as lemmas in Section 3. The proof of the upper bound $E[Q^N] = O(1/N)$ is given in Sections 4–6. The paper concludes in Section 7 with a brief discussion of extensions and open problems.

2. Prior Results

Consider the packet at the head of any given nonempty queue. Since travel times are geometrically distributed with parameter μ/N, the probability that this packet is placed on the ring in the current time step is at least μ/N; the conditional probability is precisely μ/N if the cell is occupied on arrival and it is trivially 1 if the cell is empty. Thus, one expects that, in statistical equilibrium, an individual queue length Q_i^N is bounded stochastically by the length of a single-server Markov (i.e., M/M/1) queue in discrete time with arrival and service rate parameters λ/N and μ/N. Moreover, this bound should hold independently for each queue. Indeed, these observations are but a special case of Theorem 2 in Coffman et al. (1993). An easy analysis of the discrete time M/M/1 queue then proves

Lemma 2.1. *For each i independently, Q_i^N is stochastically smaller than a non-negative integer random variable R with $P(R=n) \sim (1-\rho)\rho^n$ as $N \to \infty$ for every $n \geq 0$, and with*

$$P(R > n) = O(e^{-\nu n}) \tag{2.1}$$

where $\nu = \ln 1/\rho > 0$.

It is useful to think of the departure process as being implemented by the following mechanism. In every time step all cells independently sample a binary random variable G with $P(G=1) = \mu/N$ and $P(G=0) = 1-\mu/N$. All samples are independent of the past, so successes ($G=1$), to be called *enabling events*, form independent, identical Bernoulli processes at the N cells. An enabled, occupied cell releases/delivers its packet and then replaces it with a new packet if it is in front of a nonempty queue. Enabling events at cells that are empty and hence already enabled, have no effect. It is obvious that this mechanism for releasing packets from the ring leads to geometric transit times with parameter μ/N, as stated.

Hereafter, we take the equivalent point of view that *the queues rotate past the ring of cells, which remains fixed.* As shown in Fig. 2, in any given time interval $[0,T]$, the ring process can be represented by events on a cylindrical lattice cut at some cell position and laid out as a rectangle. For simplicity, we assume that the cylinder is cut between cell N and cell 1. Along the top of the rectangle the $Q_i^N(0)$, $1 \leq i \leq N$, give the initial state of the queues, and the crosses ($\times$'s) indicate the initial cell states: a cell with a $\times$ at time 0 is empty, otherwise, it is occupied. Again for simplicity, we assume queue 1 is at cell 1 at time 0. Within the rectangle, circles ($\circ$'s) and $\times$'s give a random sample of new arrivals and enabling events, respectively. Although not illustrated in the figure, a $\times$ and $\circ$ can appear at the same lattice point; the probability of such an event is $O(1/N^2)$ and hence relatively low.

The greedy policy is represented by a suitable matching of $\times$'s to $\circ$'s (new arrivals) and to packets in the initial state. An example is shown in Fig. 2. The broken *matching line* drawn between a matched packet and a $\times$ has a diagonal part describing the motion of the packet in time and space and a vertical part extending to a $\times$ in the cell where the packet is placed. A diagonal part is broken into two pieces when it extends past cell N, one ending

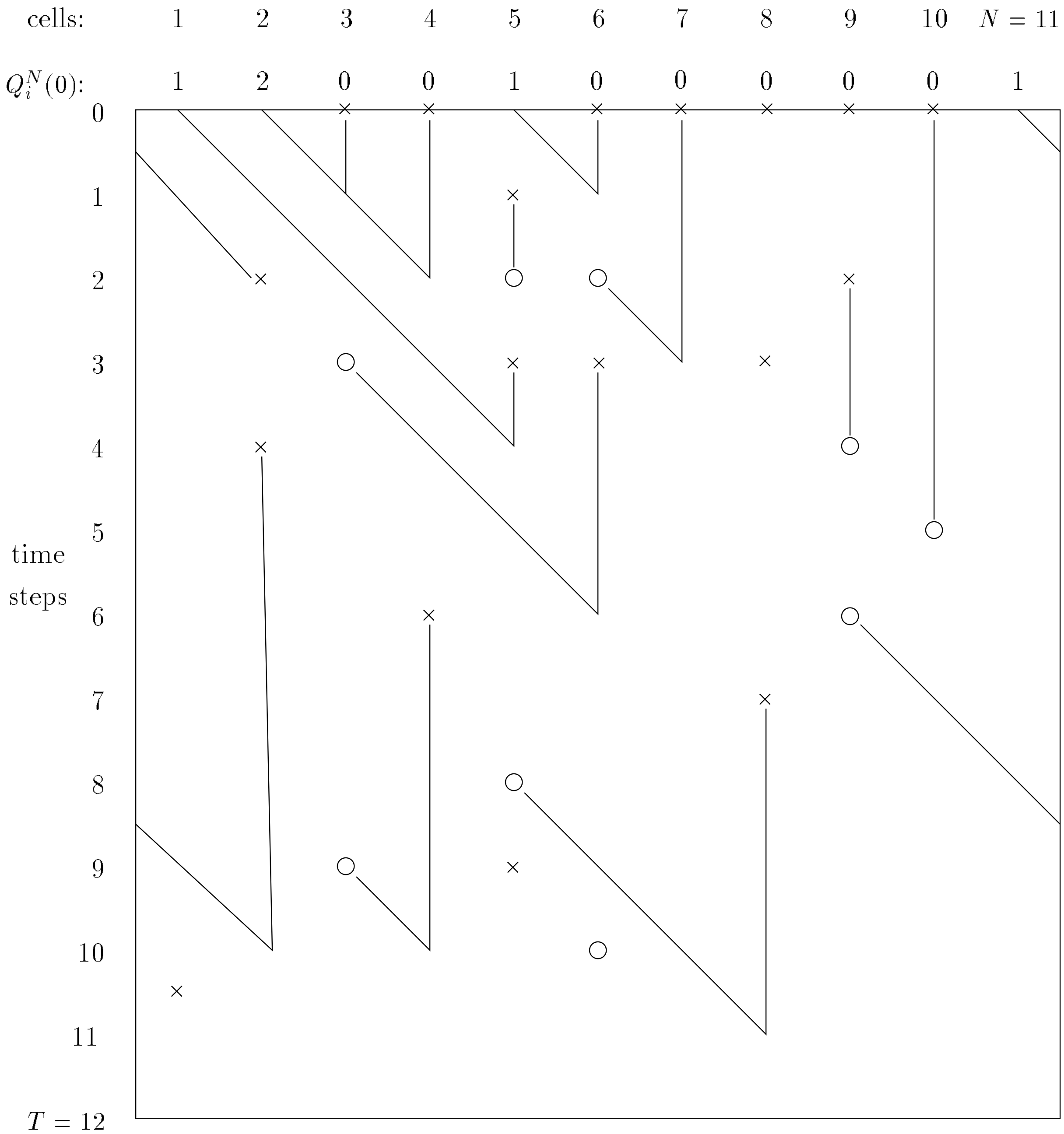

Figure 2: Greedy Matching

at the right boundary and one beginning at the same time at the left boundary.

We distinguish the vertical *part* of a matching line from its vertical *component*; the latter refers to the vertical component of the diagonal part, which has the interpretation of waiting time during $[0,T]$.

Note carefully that some $\times$'s must go unused in general. This is because the vertical part of a matching line can not pass through a $\times$. For example, the arrival at time 10 in column 6 can not be matched to either of the two unmatched $\times$'s in column 8; this would imply that the arrival is placed into cell 8 at time 12 while it is still occupied by the arrival at time 8 in column 5. This proscription does not apply to $\circ$'s that can appear on matching lines, or to $\times$'s that can appear on diagonal parts of matching lines.

Let L denote the set of enabling events over the N cells during $[0,T]$, and define $K = K_0 \cup K_a$, where K_0 is the set of packets in the initial queue state and K_a is the set of arrivals during $[0,T]$. In the following, ϕ denotes a packet and ψ denotes an enabling event.

In Fig. 2, K and L can be taken as a random instance for any of a large class of admission policies, in particular any policy that, like the greedy rule, never removes a packet from the ring except when it departs for good. Each such policy will induce a matching between some subset $K' \subseteq K$ and an equal cardinality subset $L' \subseteq L$ which specifies the packets admitted to the ring in $[0,T]$. To ensure a valid matching, we require that, if packet $\phi \in K'$ is matched to $\psi \in L'$, then

(i) the diagonal part of the matching line starting at ϕ must extend below ψ at some time no later than T,

(ii) the vertical part of the matching line terminating at ψ must pass through no enabling $\psi \neq \psi'$.

We call such policies *hot potato* policies consistent with the use of this term in studies of processor interconnection networks where packets must be kept in motion on the network until they depart once and for all (a packet can not be removed temporarily and placed in some queue along the way).

Now define the total customer waiting time in queue during $[0,T]$,

$$S(T) = \sum_{\phi \in K} w(\phi) \ , \tag{2.2}$$

where $w(\phi) = w(\phi, T)$ is the waiting time in $[0, T]$ of packet ϕ. Coffman et al. (1993) prove the following deterministic optimality result for the greedy admission policy.

Lemma 2.2. *For given sets K and L, the greedy policy has the smallest sum of waiting times $S(T)$ among the class of hot potato policies.*

3. Probability Bounds

We begin with a useful Chernoff bound that combines Theorems A.12 and A.13, pp. 237–238, in Alon and Spencer (1991).

Lemma 3.1. *Let $Z = Z_1 + \ldots + Z_n$, where the Z_i are independent Bernoulli random variables with $P(Z_i = 1) = p_i$, $P(Z_i = 0) = 1 - p_i$. Then for any $\epsilon > 0$, there exists a $\beta > 0$ such that*

$$P((1-\epsilon)E[Z] < Z < (1+\epsilon)E[Z]) = 1 - O(e^{-\beta E[Z]}) \ . \tag{3.1}$$

Next, we consider a Lindley process, starting at the origin and defined by

$$\zeta_0 = 0, \quad \zeta_i = (\zeta_{i-1} + U_i)^+ \ , \tag{3.2}$$

with $U_i = X_i - Y_i$, where $\{X_i\}$ and $\{Y_i\}$ are independent sequences of i.i.d. random variables. In our application, Y_i is a 0-1 random variable and X_i is the number of arrivals to a queue in bN time steps, where b is a given constant. Thus, for large N, X_i is approximately Poisson distributed with mean λb. It is easy to check that X_i and hence U_i has an exponential tail probability, i.e., there exists a $\kappa > 0$ such that

$$P(U_i > x) = O(e^{-\kappa x}) \ . \tag{3.3}$$

The process $\{\zeta_i\}$ is said to have negative drift if $P(Y_i = 1) = E[Y_i] > E[X_i]$ and hence $E[U_i] < 0$. The next result follows from standard theory (e.g., see Asmussen (1987)). Let the U_i be distributed as U.

Lemma 3.2. *If $E[U] < 0$, then $E[\zeta_i]$ is bounded by a constant uniformly in $i \geq 0$. The distributions of the ζ_i converge in total variation geometrically fast to the distribution of a random variable ζ with moments of all orders.*

In addition to Lemma 3.2, we will need certain probability bounds on excursions of $\{\zeta_i\}$. These will be derived in terms of corresponding bounds for the unrestricted process

$$\xi_i = \xi_{i-1} + U_i, \quad i \geq 1 , \tag{3.4}$$

with the U_i defined as before, and with a given initial state ξ_0. Hereafter, we assume a negative drift $E[U] < 0$.

The probability bound on excursions of $\{\xi_i\}$ that we will use in the analysis of $\{\zeta_i\}$ is developed as follows. Since $E[U] < 0$, and $P(U > 0) > 0$, there exists an $\alpha_0 > 0$ such that $E[e^{\alpha_0 U}] = 1$. Define the process $\xi_i^* = e^{\alpha_0 \xi_i}$, $i \geq 0$, with the property

$$E[\xi_{i+1}^* \mid \xi_i^*] = E[e^{\alpha_o U_i} \xi_i^* \mid \xi_i^*] = \xi_i^* E[e^{\alpha_0 U}] = \xi_i^* \ .$$

Together with our assumptions on U, this shows that $\{\xi_i^*\}$ is a uniformly integrable martingale, so we have

$$\begin{aligned} P\left(\sup_{i\geq 0} \xi_i \geq x\right) &= P\left(\sup_{i \geq 0} \xi_i^* \geq e^{\alpha_0 x}\right) \\ &\leq e^{-\alpha_0 x} E[\xi_0^*] = E[e^{-\alpha_0 (x - \xi_0)}] \ , \end{aligned} \tag{3.5}$$

where the inequality follows from Doob's martingale inequality (see, for example, Section 35 in Billingsley (1986)).

We now use (3.5) to get similar bounds for the *busy periods* of $\{\zeta_i\}$. In analogy with queueing applications, we say that steps i_1 through i_2, $i_2 > i_1$, comprise a busy period if $\{\zeta_i\}$ moves away from the origin at step $i_1 \geq 1$ and makes its first subsequent return to the origin at step i_2, i.e., $\zeta_{i_1 - 1} = 0$, $\zeta_j > 0$, $i_1 \leq j < i_2$, and $\zeta_{i_2} = 0$. The process is *idle* while it resides at the origin. We want a probability bound on the maximum value of the process during a busy period B. For this purpose, we make use of the fact that, away from the origin, $\{\zeta_i\}$ behaves as an unrestricted random walk. In particular, the conditional probability that, given the first jump $U_{i_1} > 0$, $\{\zeta_i\}$ exceeds level x before its next return to the origin is the same as the probability that, starting in state U_{i_1}, the unrestricted version $\{\xi_i\}$ exceeds level x before its first passage to a point at or below the origin. As an easy consequence of (3.5), we have that, for a randomly chosen busy period B of $\{\zeta_i\}$,

$$P\left(\sup_{i \in B} \zeta_i > x\right) \leq E[e^{-\alpha_0 (x - U_+)}] = O(e^{-\alpha_0 x}) \ , \tag{3.6}$$

where U_+ has the conditional distribution of U given that $U > 0$.

Our primary interest is in the behavior of $\{\zeta_i\}$ over a finite (and large) number of steps. It is convenient to let N denote the number of steps, since in later applications of the results below, N will also denote the ring size. For example, a bound on $P(\sup_{1\le i\le N}\zeta_i > \alpha \ln N)$, $\alpha > 0$, will be useful. To get such a bound, note that there are at most $N/2$ busy periods in the first N steps of $\{\zeta_i\}$. Then by (3.6)

$$P\left(\sup_{1\le i\le N}\zeta_i > x\right) \le \frac{N}{2}P\left(\sup_{i\in B}\zeta_i > x\right) = O(e^{-\alpha_0 x+\ln N}) \ .$$

Thus, for any $\gamma > 0$, we can choose $x = x(N) = \alpha \ln N$ with $\alpha = \alpha(\gamma)$ sufficiently large that

$$P\left(\sup_{1\le i\le N}\zeta_i > \alpha \ln N\right) = O(e^{-\gamma \ln N}) = O(N^{-\gamma}) \ . \tag{3.7}$$

Consider next the duration D of a randomly chosen busy period B.

Lemma 3.3. *There exists an $\eta_0 > 0$ such that*

$$P(D > y) = O(e^{-\eta_0 y}) \ .$$

Proof: Let $\{U_i\}$ be the common sequence generating both $\{\zeta_i\}$ and $\{\xi_i\}$, $\zeta_0 = \xi_0 = 0$, and suppose the first busy period B_1 of $\{\zeta_i\}$ begins at step $\ell \ge 1$. Let D_1 be the duration of B_1. It is easy to check that, for any integer $y \ge 1$, the event $\{\zeta_i > 0 \text{ for all } i, \ell \le i \le \ell + y\}$ implies the event $\{\xi_{\ell+y} \ge \xi_\ell\}$. Busy periods are i.i.d. and $P(\xi_{\ell+y} \ge \xi_\ell)$ does not depend on ℓ, so

$$\begin{aligned} P(D > y) = P(D_1 > y) &\le P(\xi_{\ell+y} \ge \xi_\ell) \\ &\le P(\xi_y \ge 0) \ , \end{aligned} \tag{3.8}$$

By Lemma 3.1, we obtain that, for any $\epsilon > 0$, there exists an $\alpha > 0$ such that

$$P(\xi_y > (1-\epsilon)E[\xi_y]) = O(e^{\alpha E[\xi_y]}) \ , \tag{3.9}$$

with $E[\xi_y] = yE[U] < 0$. To see this, we need only observe that the U_i and hence ξ_y can be expressed as sums of independent 0-1 random variables. Put $\epsilon = 1$ in (3.9) and conclude that, for some $\alpha_1 > 0$,

$$P(\xi_y \ge 0) = O(e^{\alpha_1 y E[U]}) \ . \tag{3.10}$$

Together with (3.8), this proves the lemma. ∎

We need a final observation on the age (or elapsed time) A^N of the busy period, if any, in progress at a given time step N, i.e., $A^N = N - j$ where j, $0 \le j \le N$, is the largest integer no greater than N such that $\zeta_j = 0$. For large N, the behavior of the process is approximately stationary around N. Standard estimates show that, for N large enough, A^N is stochastically less than a random variable with the elapsed-time (or residual-life) distribution $g_i = \frac{1-F_{i-1}}{E[D]}$, $i \ge 1$, where F_i is the cumulative distribution function of the duration D of a random busy period. By Lemma 3.3, $1 - F_{i-1} = O(e^{-\eta_0 i})$ so that $\sum_{i \ge n} g_i = O(e^{-\eta_0 n})$. We conclude that, for every N large enough

$$P(A^N > y) = O(e^{-\eta_0 y}) \, , \tag{3.11}$$

i.e., we obtain the same bound as in Lemma 3.3, within a constant factor.

4. Proof of Theorem 1.1: Overview

As we will be concerned with asymptotics in N, a superscript N will relate important quantities to the ring size in the remaining sections. The proof of Theorem 1.1 estimates the expectation of the sum $S^N \equiv S(T^N)$ of waiting times in an interval of length $\Theta(N^3)$, assuming that the state of the queues at the beginning of the interval is a sample from the stationary distribution. For convenience, we take $[0, T^N]$ as the interval. To make use of this estimate, observe that

$$S^N = \sum_{t=0}^{T^N} \sum_{i=1}^{N} Q_i^N(t) = \sum_{\phi \in K^N} w(\phi) \, , \tag{4.1}$$

where K^N is the set of packets with waiting times wholly or partially in $[0, T^N]$. When the system is stationary, $E[Q_i^N(t)] = E[Q^N]$, so $E[S^N] = NT^N E[Q^N]$ and

$$E[Q^N] = \frac{E[S^N]}{NT^N} \, . \tag{4.2}$$

We will prove that, under a matching admission policy to be defined in the next section, the sum of waiting times $\tilde{S}^N$ over $[0, T^N]$ satisfies $E[\tilde{S}^N] = O(N^3)$. By Lemma 2.2 $E[S^N] \le E[\tilde{S}^N]$; substitution into (4.2) then proves $E[Q^N] = O(1/N)$, since $T^N = \Theta(N^3)$.

We give a brief outline of the proof based on Fig. 3, as follows. As shown in the figure, the interval $[0, T^N]$ is divided into 3 stages. For positive constants b and c to be determined, each depending only on λ and μ, Stage 1 lasts for $T_1^N \equiv cN\ln N$ steps, Stage 2 lasts for $T_2^N \equiv bN + cN\ln N$ steps, and Stage 3 lasts for $T_3^N \equiv bN^3 - 2bN$ steps, $T^N = T_1^N + T_2^N + T_3^N$. Technically, the expressions for T_i^N (and others to come) should be enclosed in floor or ceiling notation. Since such refinements have no effect on our asymptotic analysis for large N, we omit them to avoid distracting clutter.

In the product space of time and cell index, the 3 stages are partitioned into blocks as shown in Fig. 3; this partition serves as the basis of a matching admission policy called MATCH, defined in the next section. Stage 1 consists of an $I_N \times J_N$ array of blocks H_{ij}^N, $1 \leq i \leq I_N$, $1 \leq j \leq J_N$, with $I_N = \sqrt{cN\ln N}$ and $J_N = N/I_N$, where each block is an $I_N \times I_N$ grid. The first part of Stage 2 has an identical array of blocks H_{ij}^N, $I_N + 1 \leq i \leq 2I_N$, $1 \leq j \leq N$. The second and last part of Stage 2 consists of a block H_0^N of duration bN that extends across all of the cells. Stage 3 consists of $N^2 - 2$ identical blocks H_k^N, $1 \leq k \leq N^2 - 2$, each the same as H_0^N. Note that the choice of parameters gives us the simple bound $T^N \leq bN^3$ for any waiting time in $[0, T^N]$, for all N sufficiently large.

Under MATCH, the matching in each stage will have properties that hold *with high probability.* This phrase means: with a probability $1 - O(N^{-q})$, where $q = q(b,c)$ can be made as large as desired by suitable choices for the constants $b, c > 0$ (b will have to be taken sufficiently small and c will have to be taken sufficiently large). We will prove in Section 6 that, with high probability, MATCH constructs a matching in which

(i) all packets in the initial queue state and all new arrivals in the first I_N rows of blocks H_{ij}^N, i.e., all new arrivals in $[0, T_1^N]$, are matched to enablings in $[0, T_1^N]$; and the sum of waiting times of these packets is $o(N^3)$ as $N \to \infty$.

(ii) all new arrivals in the I_N rows of blocks H_{ij}^N in Stage 2 are matched to enablings in these blocks; the sum of waiting times of these new arrivals is $o(N^3)$ as $N \to \infty$; and an unmatched enabling is left in every d^{th} column beginning with column d, where d is a constant to be determined from λ, μ, and b.

(iii) all of the new arrivals in H_0^N are matched to the enablings left in every d^{th} column by (ii) above, and have a $O(N)$ expected total waiting time.

(iv) for all k, $1 \leq k < N^2 - 2$, all of the new arrivals in H_k^N are matched to enablings in

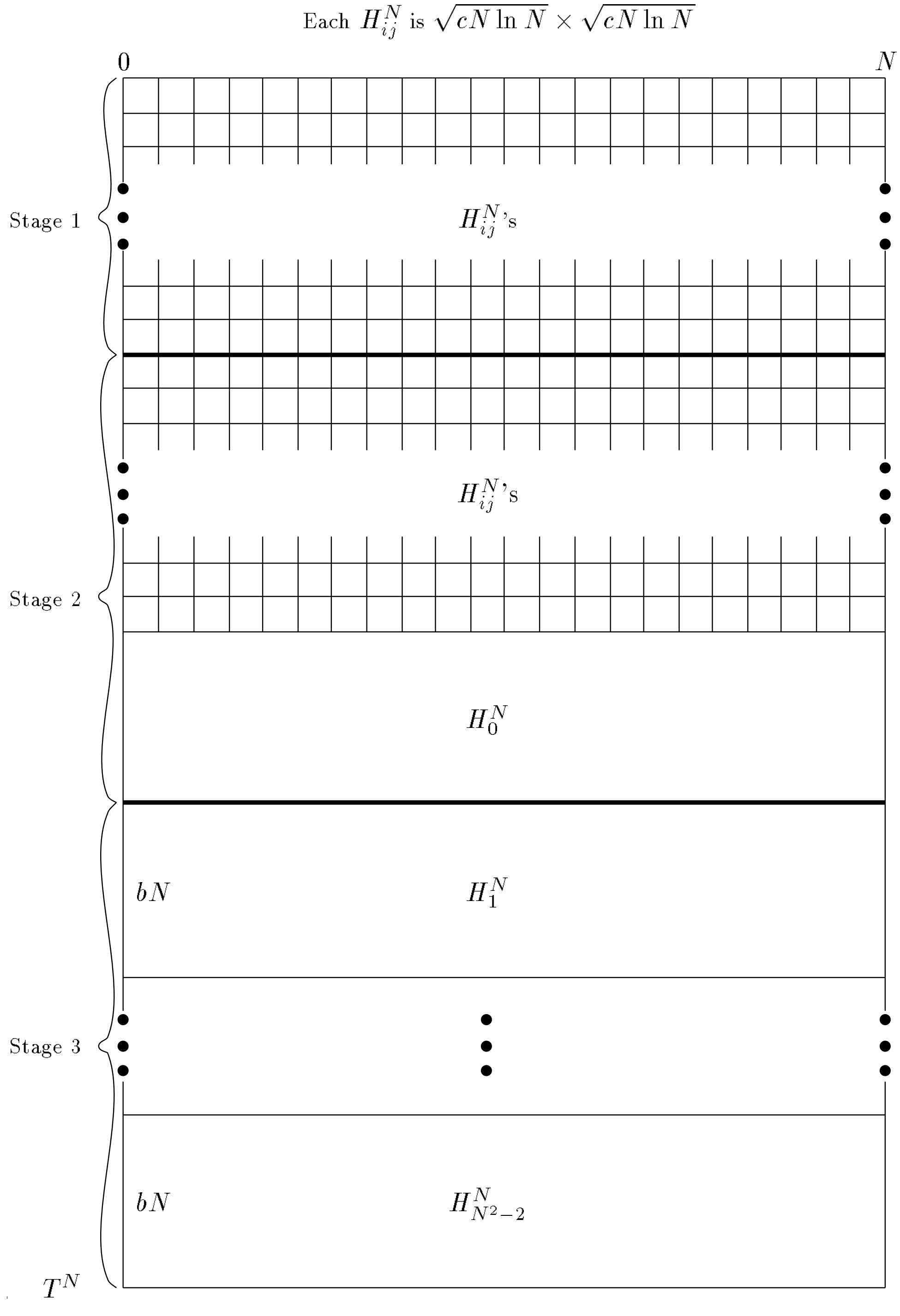

Figure 3: Partition of $[0, T^N] \times \{1, \ldots, N\}$.

H^N_{k-1} and have a $O(N)$ expected total waiting time.

The total waiting time of packets in K^N will then be $O(N^3)$ with high probability; from this fact, we will have no difficulty in proving the expected value results of Theorem 1.1.

5. Proof of Theorem 1.1: Algorithm MATCH

After defining MATCH, we will discuss the matchings it produces and verify that they are valid. MATCH consists of a *thinning phase* followed by a *matching phase*. The purpose of thinning is to remove a subset of $\times$'s in such a way as to guarantee that the matchings constructed in the matching phase will be valid; in particular, the vertical parts of matching lines do not pass through enablings (see condition (ii) in Section 2). Since the steps of the thinning phase are determined by the requirements of the matching phase, the reader may find it helpful to study the matching phase first, referring to the thinning phase as necessary to verify that matchings are valid. As part of the probabilistic analysis of MATCH in the next section, we prove that the number of $\times$'s lost in the thinning phase is negligible.

The thinning phase is given in Fig. 4. For this phase (see Step 3a) and the matching phase, the *header strip* of each H^N_k, $0 \le k \le N^2 - 2$, is defined to be the first $3\sqrt{cN \ln N}$ time steps across the N columns.

In what follows, we adopt circular indexing implicitly, i.e., when we refer to a block $H_{i,j+k}$ or $H_{i,j-k}$, $0 < k < N$, then $i + k > N$ is to be taken as $i + k - N$ and $i - k \le 0$ is to be taken as $N + (i - k)$. The matching phase defined in Fig. 5 is divided into stages corresponding to the stages of $[0, T^N]$ illustrated in Fig. 3. The algorithm refers to *corner blocks* of H^N_k, $1 \le k \le N^2 - 1$; these are small blocks in the lower left corners consisting of the last $\sqrt{cN \ln N}$ time steps and the first $\sqrt{cN \ln N}$ columns; see Fig. 8 for an illustration.

MATCH: Thinning Phase

1. (Block rows $1, \ldots, I_N + 3$ of the H_{ij}^N's)

   a. In each column segment that spans the first 3 block rows, delete all but one $\times$; the $\times$ retained is chosen at random from two or more.

   b. In any sequence, scan top-down each column segment spanned by block rows 4 to block row $I_N + 3$. Whenever a $\times$ is encountered, say in H_{ij}^N, $4 \leq i \leq I_N + 3$, leave it if there is no $\times$ directly above it in blocks H_{ij}^N, $H_{i-1,j}^N$, $H_{i-2,j}^N$, or $H_{i-3,j}^N$; otherwise, delete the $\times$.

2. (Block rows $I_N + 1$ to $2I_N$)

   a. First mark every d^{th} column beginning with column d and ending with column N/d.

   b. Thin the *unmarked* column segments of block rows $I_N + 4$ to $2I_N$ just as in 1b above, continuing 1b where it left off.

   c. Thin the *marked* column segments in block rows $I_N + 1$ to $2I_N$ by deleting all but one $\times$, if any, in each segment; the $\times$ retained is chosen at random from two or more.

3. (Blocks H_k^N, $0 \leq k \leq N^2 - 2$)

   a. Remove all $\times$'s from the header strips of every H_k^N, $0 \leq k \leq N^2 - 2$.

   b. In each column segment that spans H_0^N, delete all but one $\times$, if any; the $\times$ retained is chosen at random from two or more.

   c. Finally, for $k = 1, \ldots, N^2 - 2$, scan top-down each column segment that spans H_k^N; whenever a $\times$ is encountered, say in H_k^N, it is retained if and only if there is no $\times$ directly above it in H_k^N or H_{k-1}^N.

Figure 4: Algorithm MATCH, thinning phase.

Matching Phase

Stage 1 (block rows 1 to I_N)

a. For each i, j, $1 \le i \le I_N$, $1 \le j \le J_N$, match $\circ$'s in H_{ij}^N in any way to $\times$'s in $H_{i,j+2}^N$ until either the former or the latter are exhausted, whichever occurs first.

b. For each j, $1 \le j \le I_N$, match the packets in the queues $Q_{(j-1)I_N+1}^N, \ldots, Q_{jI_N}^N$ above H_{ij}^N to $\times$'s that remain in the respective blocks of the spiral sequence $H_{1,j+2}^N$, $H_{2,j+3}^N, \ldots, H_{I_N,j+I_N+1}^N$, i.e., for each $\ell = 1, \ldots, I_N$ match initial packets in $Q_{(j-1)I_N+\ell}^N$ to leftover $\times$'s in $H_{\ell,j+\ell+1}^N$ until the former or the latter is exhausted, whichever occurs first.

Stage 2 (block rows $I_N + 1$ to I_N and block H_0^N)

a. For the second I_N rows of blocks H_{ij}^N, perform the same matching as in 1a above, with one exception: use no $\times$ in a column whose number is a multiple of d; these are the marked columns of step 2a of the thinning phase.

b. Scan the N columns left to right in the time interval spanned by H_0^N, beginning with column 1. To each $\circ$ encountered, match the leftmost unmatched $\times$, if any, in a *marked* column segment directly above or above and to the right of the column containing the $\circ$.

c. If there are unmatched $\circ$'s leftover, match these in any way to the leftover $\times$'s, if any, in block $H_{2I_N,1}^N$ until either the former or the latter are exhausted, whichever occurs first. The matching order of $\circ$'s in the same column of H_0^N can be arbitrary.

d. Eliminate all matchings made in 2b or c whose matching lines extend below the header strip of H_1^N; the corresponding $\circ$'s are left unmatched.

Stage 3 (blocks H_k^N, $1 \le k \le N^2 - 2$)

a. For $k = 1, 2, \ldots, N^2 - 2$, scan the columns of H_k^N as in 2b, matching $\circ$'s to $\times$'s directly above or above and to the right in a column segment of H_{k-1}^N, where the $\times$ can be in *any* such segment (not just the marked ones as in 2b).

b. If there are unmatched $\circ$'s leftover, match these in any way to the unmatched $\times$'s, if any, in the corner block of H_{k-1}^N until either the former or the latter are exhausted, whichever occurs first. As before, the matching order of $\circ$'s in the same column can be random.

c. Eliminate all matchings made in 3a or b whose matching lines extend below the header strip of H_{k+1}^N, when $k < N^2 - 2$, or below T^N when $k = N^2 - 2$; the corresponding $\circ$'s are left unmatched.

Figure 5: Algorithm MATCH: Matching Phase.

With the help of Figs. 6-8, it is easy to verify that all matchings are valid. Consider Stage 1a illustrated in Fig. 6. As can be seen, a diagonal down from any $\circ$ in H_{ij}^N must pass under all $\times$'s in $H_{i,j+2}^N$. The vertical parts have lengths ranging from 0 to $4\sqrt{cN\ln N}$ and, because of the thinning process, no vertical part can pass through a $\times$ in $H_{i+2,j+2}^N$, $H_{i+3,j+2}^N$, or $H_{i+4,j+2}^N$ (all of the $\times$'s shown in the figure must be in different columns). The vertical components, or waiting times, are bounded by $3\sqrt{cN\ln N}$. A similar figure could be drawn for blocks H_{ij}^N, $H_{i,j+2}^N$, in Stage 2, but the $\times$'s matched would be confined to unmarked columns. Note also that matching lines from $\circ$'s in the last two rows of blocks H_{ij}^N can extend down into the header strip of H_0^N, but the vertical parts of such lines can not be in marked columns and they can not extend below the header strip of H_0^N.

Figure 7 illustrates the matching of initial packets, in this case, 2 packets in the queue over the third cell of H_{1j}^N to $\times$'s left over from Stage 1a in the third block of the shaded diagonal sequence. Extensions of the remarks above show that the matching lines are valid. The vertical parts range from a minimum of 0 to a maximum of $3\sqrt{cN\ln N}$ and the waiting times are bounded by $T_1^N + 2\sqrt{cN\ln N}$ (matching lines can extend down into the first two block rows of Stage 2).

An example of H_k^N, $k \geq 1$, in Stage 3 is shown in Fig. 8. Note that vertical parts are bounded by $2bN$, and that by the thinning process a $\times$ in column j in H_{k-1}^N disallows a $\times$ in column j of H_k^N, so vertical parts can not pass through $\times$'s. Note also that the diagonal parts of $\circ$'s in H_k^N matched to $\times$'s in the corner block of H_{k-1}^N wrap around from the right edge to the left edge. A similar illustration could be given for H_0^N; however, the $\times$'s matched would be restricted to the marked columns in Stage 2 extending from block row $I_N + 1$ down to the top of H_0^N.

6. Proof of Theorem 1.1: Probabilistic Analysis

It remains to prove the estimate $E[\tilde{S}^N] = O(N^3)$ of the expected total waiting times under MATCH. Let $\tilde{S}_i^N$ be the total waiting time contributed by packets in Stage i of the matching phase, $i = 1, 2, 3$. We will prove that $E[\tilde{S}_i^N] = O(N^3)$ for each i, so that $E[\tilde{S}^N] = E[\tilde{S}_1^N] + E[\tilde{S}_2^N] + E[\tilde{S}_3^N] = O(N^3)$, as desired. It will be convenient to analyze the stages in the order $1, 3, 2$.

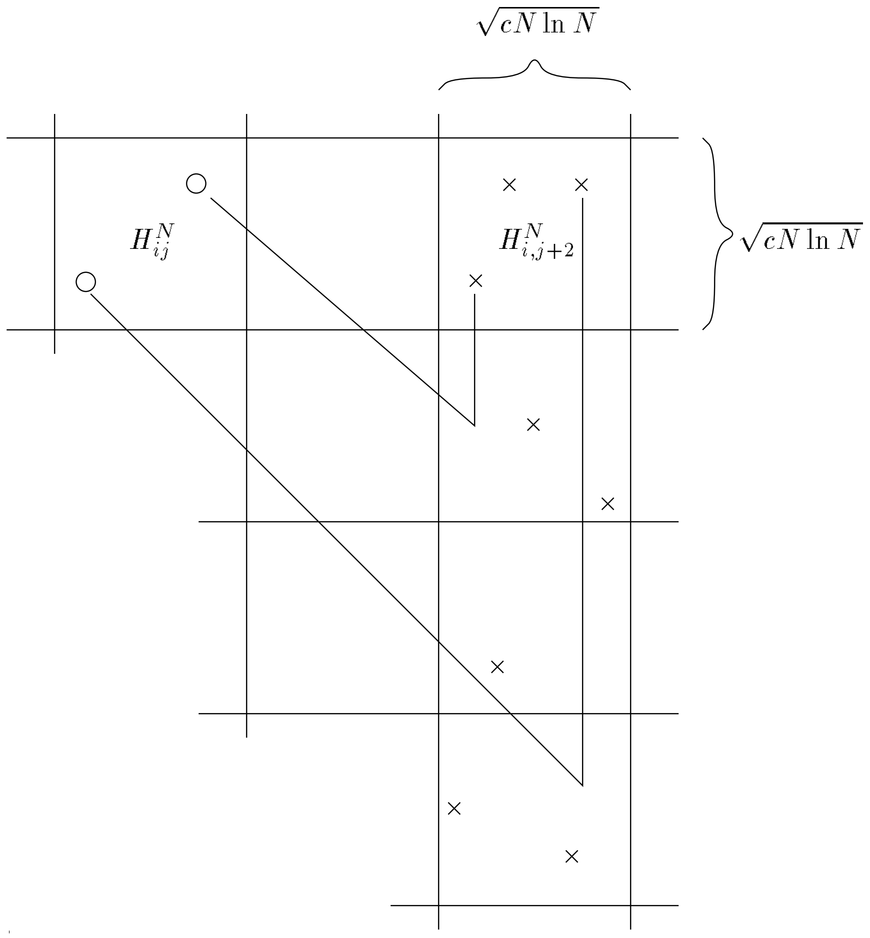

Figure 6: Matching of new arrivals in H_{ij}^N, Stage 1. Only $\circ$'s in H_{ij}^N are shown and only $\times$'s in the $(j+2)^{\mathrm{nd}}$ column of blocks are shown.

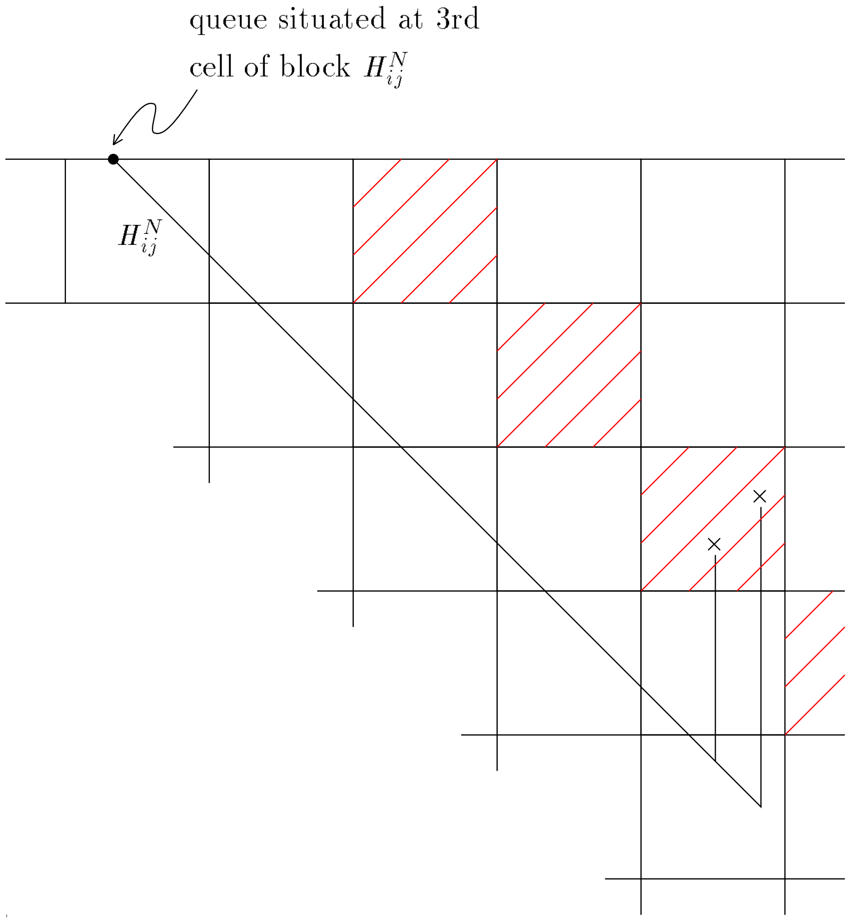

Figure 7: Matching initial packets.

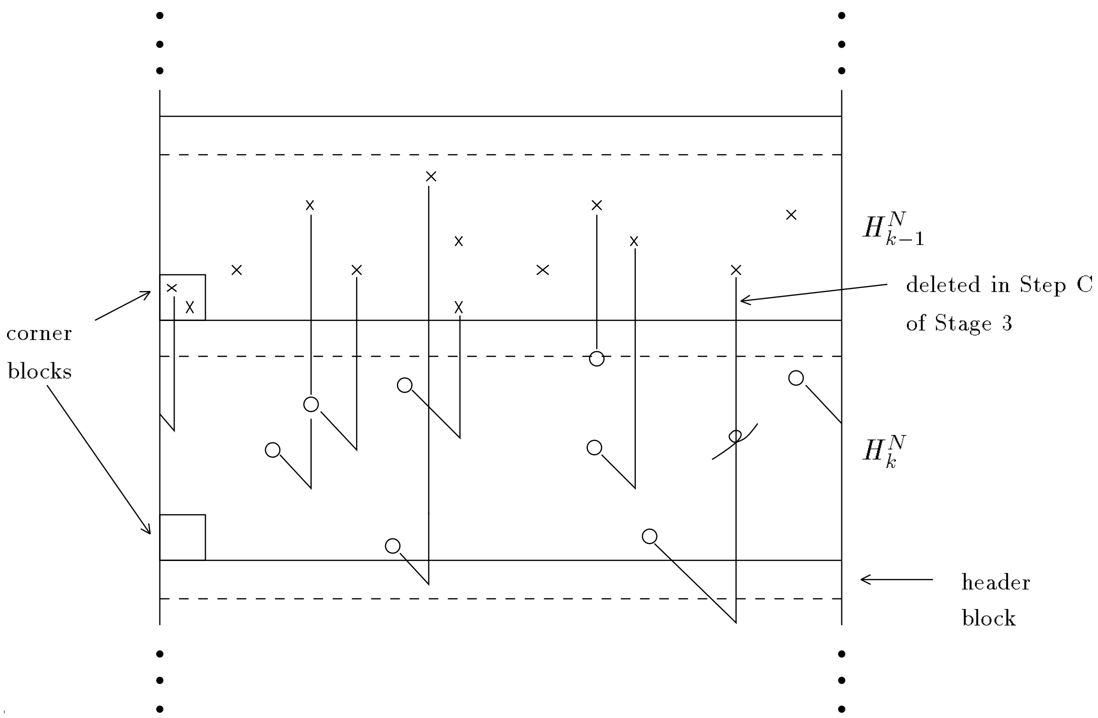

Figure 8: Matching new arrivals in H_k^N, $k \geq 1$. Only the $\circ$'s are shown in H_k^N and only the $\times$'s are shown in H_{k-1}^N.

Stage 1 Consider an arbitrary block H_{ij}^N in Stage 1. Let K_{ij}^N denote the set of new arrivals served by $\times$'s in H_{ij}^N, i.e., K_{ij}^N is the set of $\circ$'s in H_{ij-2}^N; and let Q_{ij}^N denote the set of initial packets in the queue served by H_{ij}^N. Define $n^\circ = |K_{ij}^N|$ and let $n^\times$ be the number of $\times$'s in H_{ij}^N. Our aim is to find the expected total waiting time of the packets $\phi \in K_{ij}^N \cup Q_{ij}^N$ served by H_{ij}^N.

We begin with 3 claims, the first two giving bounds on the conditional expected total waiting times of packets in Q_{ij}^N and K_{ij}^N, given the event

$$\mathcal{E}_{ij} = \{n^\times - n^\circ > \delta c \ln N\} \tag{6.1}$$

for some δ, $0 < \delta < \mu - \lambda$. The third claim shows that $\mathcal{E}_{ij}$ occurs with high probability. With these claims in hand, the remainder of the proof that $E[\tilde{S}_1^N] = O(N^3)$ will be very short. In what follows, $\tilde{w}$ denotes waiting times in $[0, T^N]$ under MATCH.

Claim 1. *For c large enough*

$$E\left[\sum_{\phi \in Q_{ij}^N} \tilde{w}(\phi) \mid \mathcal{E}_{ij}\right] = O(N \ln^2 N) \ .$$

Proof: Matched packets in Q_{ij}^N have a waiting time bounded by $cN \ln N + 2\sqrt{cN \ln N}$, which is the duration of Stage 1 plus the duration of the top two block rows of Stage 2; and unmatched packets in Q_{ij}^N have the trivial waiting-time bound $T^N \leq bN^3$. For the former we use the simpler bound $2cN \ln N$ (assuming N is large enough) and define, for any δ satisfying $0 < \delta < \mu - \lambda$,

$$\tilde{w}_m = \begin{cases} 2mcN \ln N, & 0 \leq m \leq \delta c \ln N \\ 2\delta c^2 N \ln N + (m - \delta c \ln N) bN^3, & m > \delta c \ln N \ . \end{cases} \tag{6.2}$$

Then $\mathcal{E}_{ij}$ implies (we use the independence between Q_{ij}^N and n°, $n^\times$),

$$E\left[\sum_{\phi \in Q_{ij}^N} \tilde{w}(\phi) \mid \mathcal{E}_{ij}\right] \leq \sum_{m \geq 0} \tilde{w}_m P(|Q_{ij}^N| = m) \ .$$

By Lemma 2.1, $|Q_{ij}^N|$ is bounded stochastically by R^N. Therefore, since $\tilde{w}_m$ is nondecreasing in m, we have (see Ross (1983), p. 252)

$$E\left[\sum_{\phi \in Q_{ij}^N} \tilde{w}(\phi) \mid \mathcal{E}_{ij}\right] \leq \sum_{m \geq 0} \tilde{w}_m P(R^N = m) \ .$$

Substitutions then give

$$
\begin{aligned}
E\left[\sum_{\phi\in Q_{ij}^N} \tilde{w}(\phi) \mid \mathcal{E}_{ij}\right] &\le O(N\ln^2 N) + bN^3 \sum_{m>\delta c\ln N} (m-\delta c\ln N)P(R^N = m) \\
&\le O(N\ln^2 N) + bN^3 \sum_{m>0} mP(R^N > m+\delta c\ln N)
\end{aligned} \tag{6.3}
$$

by (6.2) and a change of variables. Then (2.1) shows that

$$
P(R^N > m+\delta c\ln N) = e^{-\nu m}\cdot O(e^{-\nu\delta c\ln N}) = e^{-\nu m}\cdot O(N^{-\nu\delta c}) \ , \tag{6.4}
$$

so the sum in (6.3) is $O(N^{-\nu\delta c})\sum_{m>0} me^{-\nu m} = O(N^{-\nu\delta c})$. Then the second term on the right of (6.3) can be made as small as desired by taking c sufficiently large. The claim follows. ∎

Claim 2.

$$
E\left[\sum_{\phi\in K_{ij}^N} \tilde{w}(\phi) \mid \mathcal{E}_{ij}\right] = O(\sqrt{N}\ln^{3/2} N) \ .
$$

Proof: When $\mathcal{E}_{ij}$ holds, all new arrivals in $H_{i,j-2}^N$ are matched, and by the Stage–1 matching procedure, each has a waiting time bounded by $3\sqrt{cN\ln N}$ (see Fig. 5). There can be at most n° matched packets in K_{ij}^N, so

$$
E\left[\sum_{\phi\in K_{ij}^N} \tilde{w}(\phi) \mid \mathcal{E}_{ij}\right] \le 3\sqrt{cN\ln N}\ E[n^\circ \mid \mathcal{E}_{ij}] \ . \tag{6.5}
$$

But n° is a positive random variable, so $E[n^\circ \mid n^\circ \le x] \le E[n^\circ] = \lambda c\ln N$ for all x. Thus, $E[n^\circ \mid \mathcal{E}_{ij}] = O(\ln N)$. Substitution into (6.5) proves the claim. ∎

Claim 3. *For any $c>0$ and any δ satisfying $0<\delta<\mu-\lambda$, there exists a $\beta>0$ such that*

$$
P(n^\times - n^\circ < \delta c\ln N) = 1 - P(\mathcal{E}_{ij}) = O(N^{-\beta c}) \ .
$$

Proof: Let $n_k^\circ \in \{0,1\}$, $1\le k\le cN\ln N$, denote the number of arrivals at the k^{th} grid point of $H_{i,j-2}^N$, under any given enumeration of these points. Then $n^\circ = \sum_k n_k^\circ$. Applying Lemma 3.1 with $n = cN\ln N$ and $p_i = \lambda/N$ for all i, we have that, for any $\epsilon>0$, there exists a $\beta' = \beta'(\epsilon) > 0$ such that

$$
P(n^\circ > (1+\epsilon)\lambda c\ln N) = O(e^{-\beta'\lambda c\ln N}) = O(N^{-\beta'\lambda c}) \ . \tag{6.6}
$$

A similar bound can be obtained for $n^\times$, but because of the thinning procedure we count the $\times$'s differently. Let $n_k^\times$, $1 \le k \le \sqrt{cN \ln N}$, denote the number of $\times$'s in the $k^{\rm th}$ column segment of H_{ij}^N, so that $n^\times = \sum_k n_k^\times$. By the thinning procedure, the $n_k^\times$ are independent 0-1 random variables. Let $i > 3$ for simplicity; the arguments for $i = 1, 2, 3$ require trivial modifications. If at the beginning of the thinning procedure $H_{i-3,j}^N$, $H_{i-2,j}^N$, and $H_{i-1,j}^N$ have no $\times$'s in the $k^{\rm th}$ column segment of H_{ij}^N, but H_{ij}^N has at least one, then H_{ij}^N will have one $\times$ in its $k^{\rm th}$ column segment at the end of the thinning procedure. Thus,

$$P(n_k^\times = 1) \ge (1 - \mu/N)^{3\sqrt{cN \ln N}} \left[1 - (1 - \mu/N)^{\sqrt{cN \ln N}}\right] ,$$

so

$$P(n_k^\times = 1) \ge \mu \sqrt{\frac{c \ln N}{N}} + O\left(\frac{\ln N}{N}\right) . \tag{6.7}$$

Now use Lemma 3.1 with $n = \sqrt{cN \ln N}$ and p_i given by (6.7) for all i to obtain, for some $\beta'' = \beta''(\epsilon) > 0$,

$$P(n^\times < (1 - \epsilon)\mu c \ln N) = O(N^{-\beta'' \mu c}) . \tag{6.8}$$

Finally, choose ϵ so that $\delta = (1 - \epsilon)\mu - (1 + \epsilon)\lambda < \mu - \lambda$, noting that $\delta > 0$ requires $\epsilon < \frac{\mu - \lambda}{\mu + \lambda}$. Then observe that $n^\times - n^\circ < \delta c \ln N$ implies that either $n^\times < (1 - \epsilon)\mu c \ln N$ or $n^\circ > (1 + \epsilon)\lambda c \ln N$. We conclude that

$$P(n^\times - n^\circ < \delta c \ln N) \le O(N^{-\beta' \lambda c}) + O(N^{-\beta'' \mu c}) ,$$

so the claim follows with $\beta = \min(\lambda\beta', \mu\beta'') > 0$. ∎

There is a trivial, deterministic bound $bcN^4 \ln N$ on the total waiting time of packets in K_{ij}^N, which is obtained from the product of the number $cN \ln N$ of lattice points (potential arrival locations) and the bN^3 bound on the waiting time of any packet in $[0, T^N]$. Also, even if all initial packets go unmatched, their expected total waiting time is bounded by $bN^3 E|Q_{ij}^N| = O(N^3)$, by Lemma 2.1. Noting that the bound of Claim 2 for $\phi \in K_{ij}^N$ is negligible compared to the bound for $\phi \in Q_{ij}^N$ in Claim 1, we conclude that

$$E\left[\sum_{\phi \in Q_{ij}^N \cup K_{ij}^N} \tilde{w}(\phi)\right] = O(N \ln^2 N) P(\mathcal{E}_{ij}) + O(N^4 \ln N)[1 - P(\mathcal{E}_{ij})] . \tag{6.9}$$

Claim 3 shows that if we choose c large enough, then the second term on the right of (6.9) will be negligible by comparison with the first, so that

$$E\left[\sum_{\phi\in Q_{ij}^N\cup K_{ij}^N}\tilde{w}(\phi)\right]=O(N\ln^2 N)\ .$$

Finally, since there are N blocks H_{ij}^N in Stage 1, we obtain $E[\tilde{S}_1^N]=O(N^3)$ as desired.

Stage 3 Let K_k^N denote the set of $\circ$'s in H_k^N and define M_k^N to be the subset of these packets that become matched in Stage 3. For simplicity, we will ignore $H_{N^2-2}^N$ completely, since unlike the other Stage 3 blocks, $H_{N^2-2}^N$ it is not followed by a header strip. We lose nothing by this, since

$$E\left[\sum_{\phi\in K_{N^2-2}^N}\tilde{w}(\phi)\right]=O(N^2) \tag{6.10}$$

follows from the λbN bound on the expected number of arrivals in $H_{N^2-2}^N$ and the bN bound on the waiting time of each.

Our approach will be to add bounds on the expected total waiting times, computed separately, for the matched packets and the unmatched packets. Claim 4 below shows that, for b sufficiently small and c sufficiently large,

$$E\left[\sum_{\phi\in M_k^N}\tilde{w}(\phi)\right]=O(N),\quad 1\le k\le N^2-3\ .$$

Then Claim 5 will show that for any $\gamma>0$ we can take c large enough so that the probability of the event $\mathcal{E}_k$ that the Stage-3 matching procedure leaves one or more $\circ$'s in K_k^N unmatched is $O(N^{-\gamma})$. The number of grid points in H_k^N is bN^2, so with the bN^3 bound on any waiting time, we have a trivial b^2N^5 bound on the total waiting time of packets in K_k^N. Together with Claims 4 and 5, this shows that

$$E\left[\sum_{\phi\in K_k^N}\tilde{w}(\phi)\right]=O(N)+b^2N^5P(\mathcal{E}_k)=O(N)\ . \tag{6.11}$$

Multiplying by the number N^2-3 of blocks, we conclude that

$$E[\tilde{S}_2^N]=O(N^3)\ ,$$

as desired.

Claim 4. *Choose b so that $0 < b < \frac{2}{3}\frac{\mu-\lambda}{\mu^2}$. Then there is a c sufficiently large that*

$$E\left[\sum_{\phi\in M_k^N} \tilde{w}(\phi)\right] = O(N), \quad k = 1,\ldots,N^2-3 .$$

Proof: Let $U_i = X_i - Y_i$, $1 \le i \le N$, where X_i and Y_i are the respective numbers of $\circ$'s and $\times$'s in the i^{th} column segment of H_k^N and H_{k-1}^N. Note that $\{X_i\}$ and $\{Y_i\}$ are independent sequences of i.i.d. random variables; X_i is binomially distributed with the success (arrival) probability λ/N and number of trials bN, and, by the thinning phase, Y_i is a 0-1 random variable. The Lindley process induced by $\{X_i\}$ and $\{Y_i\}$ and the left-to-right scan of the columns in the Stage-3 matching is given by the Lindley process in Section 3,

$$\zeta_0^N = 0, \quad \zeta_i^N = (\zeta_{i-1}^N + U_i)^+, \quad i \ge 1 . \tag{6.12}$$

It is easy to see that, among the $\circ$'s scanned during Stage 3 in columns $1,\ldots,i$, ζ_i^N gives the number as yet unmatched. Thus, if m° denotes the number of $\circ$'s in H_k^N, then $m^\circ - \zeta_N^N$ counts the $\circ$'s matched in the scan of the N columns, and ζ_N^N counts the leftover $\circ$'s that are matched, to the extent possible, to $\times$'s in the corner block of H_{k-1}^N. An easy induction establishes that the sum of the horizontal components of the matching lines incident to the first $m^\circ - \zeta_N^N$ matched $\circ$'s is at most $\zeta_1^N + \ldots + \zeta_N^N$ with equality if $\zeta_N^N = 0$. (Observation (4.1) is essentially equivalent.) For the leftover $\circ$'s that become matched, N is a trivial bound on the horizontal components of their matching lines. Thus, since the horizontal component of a packet's matching line is equal to its waiting time, we conclude that

$$\sum_{\phi\in M_k^N} \tilde{w}(\phi) \le \sum_{i=1}^{N} \zeta_i^N + N\zeta_N^N . \tag{6.13}$$

We now verify that $\{\zeta_i^N\}$ has negative drift so that, by definition of $\{X_i\}$ and $\{Y_i\}$, we can apply Lemma 3.2 to show that $E[\zeta_i^N] = O(1)$ for all i; taking expected values in (6.13) then proves the claim.

To prove the negative drift $E[U_i] < 0$, we verify that, in spite of the $\times$'s lost by thinning, we can ensure that $P(Y_i = 1) = E[Y_i] > E[X_i]$. We consider the case $k > 1$; the argument for H_1^N is even simpler and hence omitted. It is enough to observe that a single $\times$ in column i of H_{k-1}^N is retained if the original sample in H_{k-1}^N had at least one $\times$ in column i outside of the header strip, and if the original sample in column i of H_{k-2}^N had no $\times$'s at all. Then

$$\begin{aligned} P(Y_i = 1) &\geq [1-(1-\mu/N)^{bN-3\sqrt{cN\ln N}}](1-\mu/N)^{bN} \\ &\sim (1-e^{-\mu b})e^{-\mu b}, \quad \text{as} \quad N \to \infty \ . \end{aligned} \tag{6.14}$$

Simple estimates show that for this to exceed $E[X_i] = \lambda b$, it is enough to require that $0 < b < \frac{2}{3}\frac{\mu-\lambda}{\mu^2}$, as stated in the claim. ∎

It remains to prove that $P(\mathcal{E})$ is sufficiently small, where $\mathcal{E}$ is the event that one or more new arrivals in H_k^N are left unmatched. We simplified the definition of MATCH by making all header strips have height $3\sqrt{cN\ln N}$. It will be seen that the result below would also hold if the heights of the header strips of H_k^N, $k \geq 1$, were taken to be $\Theta(\log N)$.

Claim 5. *For any $\gamma > 0$, c can be chosen large enough so that*

$$P(\mathcal{E}) = O(N^{-\gamma}) \ .$$

Proof: If $\mathcal{E}$ occurs then at least one of the following events occurs.

$\mathcal{E}_1$: ζ_N^N exceeds the number of $\times$'s left unmatched in the corner block of H_{k-1}^N.

$\mathcal{E}_2$: a waiting time of at least one of the first $m^\circ - \zeta_N^N$ matched $\circ$'s extends below the header strip of H_{k+1}^N.

$\mathcal{E}_3$: the waiting time of at least one matched leftover $\circ$ extends below the header strip of H_{k+1}^N.

We will verify that the claim applies to $\mathcal{E}_1$, $\mathcal{E}_2$, and $\mathcal{E}_3$ individually. Then, since $P(\mathcal{E}) \leq P(\mathcal{E}_1) + P(\mathcal{E}_2) + P(\mathcal{E}_3)$, the claim will be proved.

($\mathcal{E}_1$). Fix $\gamma > 0$. The bound in (3.7) shows that we can choose an $\alpha = \alpha(\gamma)$ large enough so that

$$P(\zeta_N^N > \alpha \ln N) \leq P\left(\sup_{1\leq i\leq N} \zeta_i^N > \alpha \ln N\right) = O(N^{-\gamma}) \ .$$

A comparison of the Stage 1 and 3 matching procedures shows easily that the number $\ell^\times$ of $\times$'s leftover in the corner block of H_{k-1}^N after the matching of H_{k-1}^N is stochastically at least as large as the random variable $n^\times - n^\circ$ in Claim 3. Then by Claim 3

$$P(\ell^\times < \delta c \ln N) \leq P(n^\times - n^\circ < \delta c \ln N) = O(N^{-\beta c}) \ ,$$

where δ has been chosen so that $0 < \delta < \mu - \lambda$, and where $\beta = \beta(\delta) > 0$. Now choose c so that both $\beta c > \gamma$ and $\delta c \geq \alpha$ hold. Then

$$\begin{aligned} P(\ell^\times < \alpha \ln N) &\leq P(n^\times - n^\circ < \alpha \ln N) \leq P(n^\times - n^\circ < \delta c \ln N) \\ &= O(N^{-\beta c}) = O(N^{-\gamma}) \ . \end{aligned}$$

It remains only to observe that $\mathcal{E}_1$ implies that either $\zeta_N^N > \alpha \ln N$ or $n^\times - n^\circ < \alpha \ln N$. Then

$$P(\mathcal{E}_1) \leq P(\zeta_N^N > \alpha \ln N) + P(\ell^\times < \alpha \ln N) = O(N^{-\gamma}) \ .$$

($\mathcal{E}_2$). For $\mathcal{E}_2$ to occur, a vertical component and hence a horizontal component of some matching line must exceed $3\sqrt{cN \ln N}$ which in turn implies that some busy period of $\{\zeta_i^N\}$ must have a length $D > 3\sqrt{cN \ln N}$. But by Lemma 3.3 there exists an $\eta_0 > 0$ such that this event has probability $O(e^{-\eta_0\sqrt{N \ln N}}) = O(N^{-\gamma})$ for all $\gamma > 0$.

($\mathcal{E}_3$). The occurrence of $\mathcal{E}_3$ implies that $\zeta_N^N > 0$ and the last, partial busy period of $\{\zeta_i^N\}$ begins at step $N - j$ for some $j > 2\sqrt{cN \ln N}$ (see Fig. 9). But by (3.11), there exists an $\eta_0 > 0$ such that the probability that the age of the partial busy period exceeds $\sqrt{cN \ln N}$ is at most $O(e^{-\eta_0\sqrt{N \ln N}}) = O(N^{-\gamma})$ for all $\gamma > 0$. The claim is thus proved. ∎

Stage 2 The analysis of Stage 2 uses precisely the same arguments as those used in the analysis of Stages 1 and 3. We will show how to adapt the earlier results to Stage 2, omitting details that are by now routine.

The calculation of expected waiting times for packets in H_{ij}^N, $I_N + 1 \leq i \leq 2I_N$, $1 \leq j \leq J_N$, proceeds as in Stage 1, but takes into account the smaller rate of $\times$'s owing to the unavailability of those in marked columns, i.e., every d^{th} column beginning with column d. The new rate of $\times$'s over all N columns which are available for the $\circ$'s in blocks H_{ij}^N is now $\mu' = \mu(1 - 1/d)$. Since this must still exceed the arrival rate λ of $\circ$'s over the N columns, we require that d satisfy the lower bound

$$d > \frac{\mu}{\mu - \lambda} \ . \tag{6.15}$$

Claim 3 can be used as is for the Stage-2 analysis, if we replace μ by μ'. To see this, consider the probability bound for $n^\times$ in the proof of Claim 3. Define $n_k^\times = 0$ if k is a

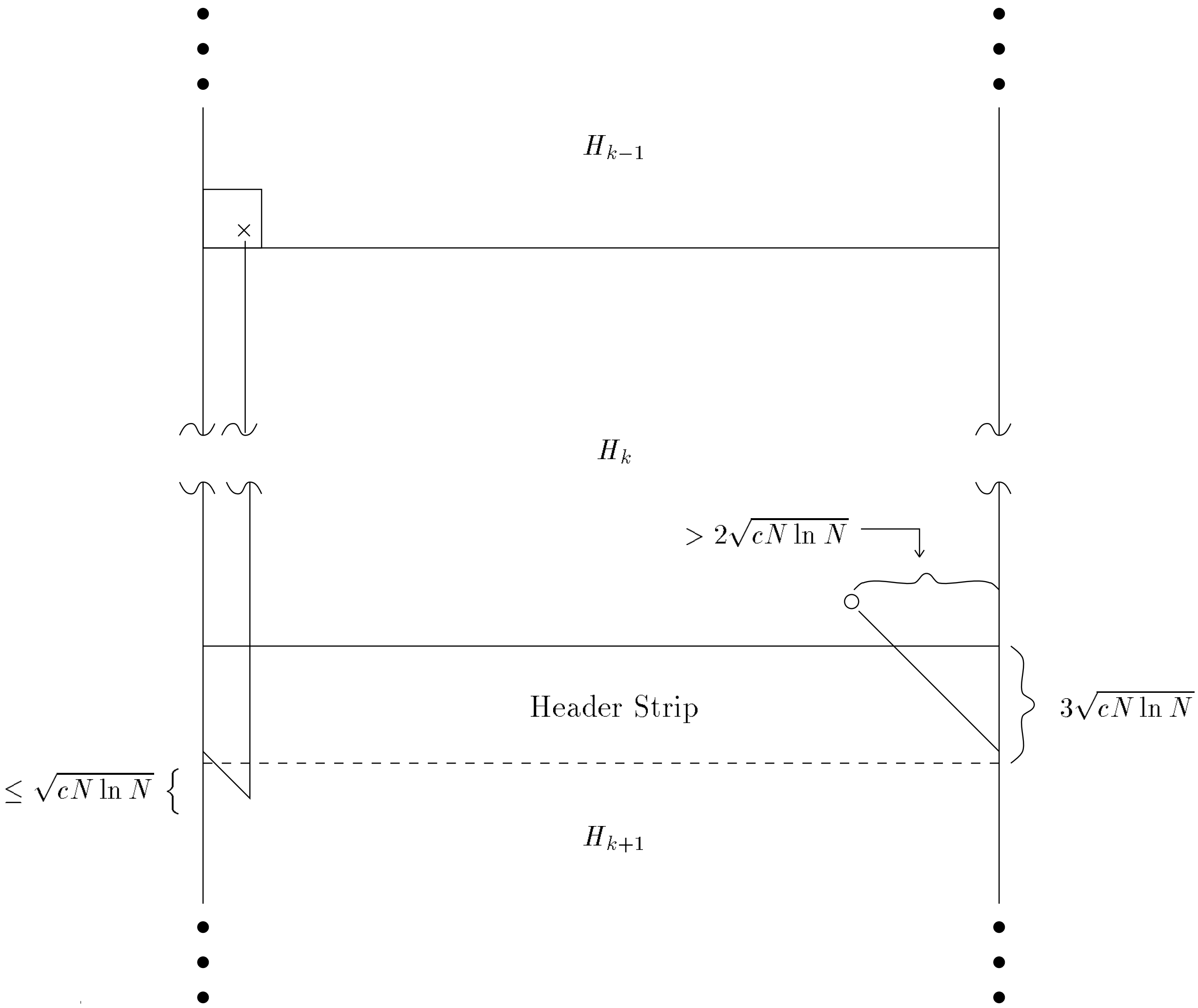

Figure 9: Illustration for $\mathcal{E}_3$; the new arrival is in the last, partial busy period.

multiple of d, so that $n^{\times}$ now counts the available $\times$'s in unmarked columns. The bound in (6.7) still holds, but in applying Lemma 3.1, we must use $n = (1 - 1/d)\sqrt{cN \ln N}$; this yields (6.8) with μ replaced by μ'. The remainder of the argument in the proof of Claim 3 holds as is with μ' replacing μ.

With the new interpretation of $n^{\times}$ in the event $\mathcal{E}_{ij}$, Claim 2 also holds as is. Thus adapted, Claims 2 and 3 then give

$$\begin{aligned} E\left[\sum_{\phi \in K_{ij}^N} \tilde{w}(\phi)\right] &= O(\sqrt{N} \ln^{3/2} N) P(\mathcal{E}_{ij}) + O(N^4 \ln N)[1 - P(\mathcal{E}_{ij})] \\ &= O(\sqrt{N} \ln^{3/2} N) \end{aligned}$$

in analogy with (6.9). Thus, the expected total waiting time of packets in block rows $I_N + 1$ to $2I_N$ is at most $O(N^{3/2} \ln N)$, which is at most $O(N^3)$ as desired.

It remains to verify a $O(N)$ bound on the expected total waiting time of packets in H_0^N. In fact, all we need is a $O(N^3)$ bound, which could be proved with cruder estimates than those used in the Stage-3 analysis. However, it is convenient to stick with the tools already in hand, thus proving the stronger result.

Since the marked columns must supply enough $\times$'s to match the $\circ$'s in H_0^N, d must satisfy an upper bound as well as the lower bound in (6.15). Before giving this bound, we discuss the Lindley process

$$\tilde{\zeta}_0^N = 0, \quad \tilde{\zeta}_j^N = \tilde{\zeta}_{j-1}^N + \tilde{U}_j, \quad j \geq 1 ,$$

constructed in analogy with the process $\{\zeta_i^N\}$ of the Stage-3 analysis. Here, only $\times$'s in marked columns are available for matching, so the epochs of $\{\tilde{\zeta}_j^N\}$ occur at every d^{th} column ending with column N, assuming for simplicity that N is a multiple of d. We have $\tilde{U}_j = \tilde{X}_j - \tilde{Y}_j$, $1 \leq j \leq N/d$, where $\tilde{X}_j$ is the number of $\circ$'s in columns $j(d-1)+1, \ldots, jd$, and $\tilde{Y}_j$ is the number of $\times$'s (0 or 1) in column jd, $j = 1, \ldots, N/d$. As before, $\tilde{\zeta}_j^N$ denotes the number of unmatched $\circ$'s that remain after scanning the first jd columns. The parameter d becomes a scale factor for the waiting times of $\circ$'s matched during the scanning process, and we obtain

$$\sum_{\phi \in M_0^N} \tilde{w}(\phi) \leq d \sum_{j=1}^{N/d} \tilde{\zeta}_j^N + N \tilde{\zeta}_N^N \tag{6.16}$$

in analogy with (6.13), where M_0^N is the set of $\circ$'s in H_0^N that are matched.

To ensure that $\{\tilde{\zeta}_j^N\}$ has negative drift, observe first that the expected number of $\circ$'s counted by $\tilde{X}_j$ is $d\lambda b$. Next, the expected number of $\times$'s counted by $\tilde{Y}_j$ is at least the probability of the event $\mathcal{E}'$ that, in the original sample of $\times$'s in the j^{th} marked column of Stage 2, at least one $\times$ occurs in the interval from $T_1^N + 3\sqrt{cN \ln N}$ to the start of H_0^N, i.e., $2T_1^N$. (Note that $\times$'s may have been deleted in the marked columns of $[T_1^N, T_1^N + 3\sqrt{cN \ln N}]$ so as to allow for matching lines extending down from $\circ$'s in blocks H_{ij}^N of rows $2I_N - 1$ and $2I_N$ in Stage 1.) Thus,

$$P(\mathcal{E}') \geq 1 - (1 - \mu/N)^{cN \ln N - 3\sqrt{cN \ln N}} \sim 1 - N^{-c\mu}, \quad N \to \infty \ ,$$

which can be made as close to 1 as desired by choosing c sufficiently large. Thus, with high probability, at the start of H_0^N, all marked columns have a $\times$ available for matching. For our upper bound on d, we require that $d\lambda b < 1$, which together with (6.15) gives

$$\frac{\mu}{\mu - \lambda} < d < \frac{1}{\lambda b} \ . \tag{6.17}$$

Then for $b < (\mu - \lambda)/\lambda\mu$, which is already assured by the choice in Claim 4, and for c sufficiently large, we can choose d to satisfy (6.17) and give the desired negative drift $E[\tilde{U}_j] < 0$. As before we take expected values in (6.16) and apply Lemma 3.2 to obtain

$$E\left[\sum_{\phi \in M_0^N} \tilde{w}(\phi)\right] = O(N)$$

Our final observation is that the proof of Claim 5 carries over directly to the analysis of $\{\tilde{\zeta}_j^N\}$ and shows that for any $\gamma > 0$ we can choose c large enough such that the probability $P(\mathcal{E})$ that a packet in H_0^N is left unmatched is $O(N^{-\gamma})$. We note a minor tightening of the argument proving $P(\mathcal{E}_1) = O(N^{-\gamma})$, viz., that $\ell^\times$ is stochastically *equal to* the random variable $n^\times - n^\circ$ in Claim 3.

This completes the proof of Theorem 1.1. ∎

7. Final Remarks

The hidden multiplicative constants in the results of Theorem 1.1 depend on λ and μ. However, a closer look at the analysis in Section 6 will provide bounds as functions of both

N and $\rho = \lambda/\mu$. In particular, one can show that there exists a universal constant α such that for all N large enough

$$E[W^N] \le \frac{\alpha}{(1-\rho)^2} \ . \tag{7.1}$$

We sketch the proof of (7.1) below; it lacks only elementary estimates from being complete.

Note first that, by (4.2) and Lemma 2.2,

$$E[Q^N] \le \frac{E[\tilde{S}^N]}{NT_N} = \frac{\lambda \overline{w}^N}{N}(1+o(1)) \quad \text{as} \quad N \to \infty \ ,$$

where $\overline{w}^N$ is the average of $\tilde{w}(\phi)$ over all packets ϕ matched in $[0, T^N]$. Thus,

$$E[W^N] \le \overline{w}^N (1+o(1)) \quad \text{as} \quad N \to \infty \ . \tag{7.2}$$

But the proof of Theorem 1.1 shows that $\overline{w}^N$ is dominated by the average of $\tilde{w}(\phi)$ for $\phi \in M_k^N$, i.e., the ϕ matched in the random walk of Stage 3. Associating the random walk $\{\zeta_i^N\}$ with a queueing process, we see that $\overline{w}^N$ is at most the waiting time in a discrete-time $G/G/1$ queue, where the arrivals in each time slot are independent and have a binomial distribution with parameters $\left(\frac{\lambda}{N}, bN\right)$, and where the service time is geometric with parameter $\mu' \approx (1-e^{-\mu b})e^{-\mu b}$, by (6.14). Simple estimates show that we can approximate this queue by an $M/G/1$ queue and deduce that (Kleinrock (1975), Section 5.7)

$$\overline{w}^N = O\left(\frac{1}{(1-\rho')\mu'}\right) \ , \tag{7.3}$$

where $\rho' = \lambda'/\mu'$ and $\lambda' = \lambda b$.

By the proof of Claim 4, if $b = \epsilon \frac{\mu - \lambda}{\mu^2}$, where $\epsilon > 0$ is a sufficiently small constant, it is readily verified that $1 - \rho' = \Omega(1-\rho)$ and $\mu' = \Omega(\mu b) = \Omega(1-\rho)$. Thus, $\overline{w}^N = O\left(\frac{1}{(1-\rho)^2}\right)$ by (7.3), as desired. Note that, unlike classical queueing systems, the expected waiting time is upper bounded by a function only of ρ, independently of the individual values of λ and μ.

There are a number of intriguing open problems in the analysis of ring communications. For example, by generalizing the transit-time distribution, we have problems of two types: stability questions and asymptotics in the ring size, N. Coffman, et al. (1993) show that, for the greedy rule in our model, the necessary condition

$$\frac{\lambda}{N} \ E[\text{transit time}] < 1 \tag{7.4}$$

is also sufficient for stability if all transit times are 1, if all transit times are N, or if the transit times are geometric with parameter μ/N. But whether (7.4) is sufficient for any other transit-time distribution is an open question.

Asymptotics in N also pose open problems for transit-time distributions other than the geometric. The uniform distribution on $\{1, \ldots, N-1\}$ is of particular interest; extensive simulations by Coffman et al. (1993) give convincing evidence that the bounds in Theorem 1.1 hold for this case as well, but no proof has yet been found.

Finally, keeping with our Markov arrival and transit-time assumptions, it would be interesting to study asymptotic behavior in the generalization of rings to toroidal arrays of processors (see Leighton (1990, 1992)). Much is known about regular (open) arrays, as can be seen from the recent work of Mitzenmacher (1994), who gives references to the earlier work on this problem. But the analysis of toroidal arrays seems to require different methods.

Acknowledgment

We are grateful to I. Telatar and A. Weiss for helpful discussions.

References

Alon, N. and Spencer, J. H. (1991), *The Probabilistic Method*, Wiley & Sons, New York.

Asmussen, S. (1987), *Applied Probability and Queues*, Wiley & Sons, New York.

Barroso, L. A. and Dubois, M. (1992), "The Performance of Cache-Coherent Ring-Based Multiprocessors," *Proc. 20th Ann. Internat. ACM Symp. Comp. Arch.*, 268–277.

Billingsley, P. (1986), *Probability and Measure*, 2nd edition, Wiley & Sons, New York.

Coffman, E. G., Jr., Gilbert, E. N., Greenberg, A. G., Leighton, F. T., Robert, P. and Stolyar, A. L., "Queues Served by a Rotating Ring," AT&T Bell Laboratories, Murray Hill, NJ 07974 (to appear).

Georgiadis, L., Szpankowski, W., and Tassiulas, L. (1994), "A Scheduling Policy with Maximal Stability Region for Ring Networks with Spatial Reuse," preprint. See also "Stability Analysis of Scheduling Policies in Ring Networks with Spatial Reuse," *Proc. 31st Ann. Allerton Conf. Comm. Cont. Comp.*, University of Illinois, Urbana.

Kleinrock, L. (1975), *Queueing Systems, Vol. I*, Wiley & Sons, New York.

Leighton, F. T. (1990), "Average Case Analysis of Greedy Routing Algorithms on Arrays," *Proc. 2nd Ann. ACM Symp. Parallel Algs. Arch.*, 2–10.

Leighton, F. T. (1992), *Introduction to Parallel Algorithms and Architectures: Arrays, Trees, Hypercubes*, Morgan Kaufmann, San Mateo, CA.

Mitzenmacher, M. (1994), "Bounds on the Greedy Algorithm for Array Networks," *Proc. 26th Ann. Symp. Th. Comp.* (to appear).

Ross, S. (1983), *Stochastic Processes*, Wiley & Sons, New York.

Van Arem, B. and Van Doorn, E. A. (1990), "Analysis of a Queueing Model for Slotted Ring Networks," *Computer Networks and ISDN Systems*, **20**, 309–314.