1990

Maximum Size of a Dynamic Data Structure: Hashing with Lazy Delection Revisited

David Aldous

Micha Hofri

Wojciech Szpankowski
Purdue University, spa@cs.purdue.edu

Report Number:
90-992

MAXIMUM SIZE OF A DYNAMIC
DATE STRUCTURE

David Aldous
Micha Hofri
Wojciech Szpankowski

CSD-TR-992
July 1990
(Revised April 1992)

MAXIMUM SIZE OF A DYNAMIC DATA STRUCTURE:

Hashing with Lazy Deletion Revisited

March 20, 1991

David Aldous*
Department of Statistics
University of California
Berkeley, CA 94720
U.S.A.

Micha Hofri†
Dept. of Computer Science
The Technion
32000 Haifa
Israel

Wojciech Szpankowski‡
Dept. of Computer Science
Purdue University
W. Lafayette, IN 47907
U.S.A.

Abstract

We study the dynamic data structure management technique called *Hashing with Lazy Deletion* (HwLD). A table managed under HwLD is built via a sequence of insertions and deletions of items. When hashing with lazy deletions, one does not delete items as soon as possible, but keeps more items in the data structure than *immediate-deletion* strategies would. This deferral allows the use of a simpler deletion algorithm, leading to a lower overhead—in space and time—for the HwLD implementation. It is of interest to know how much extra space is used by HwLD. We investigate the maximum size and the excess space used by HwLD, under general probabilistic assumptions, using the methodology of queueing theory. In particular, we find that for the Poisson arrivals and general life-time distribution of items, the excess space does not exceed the number of buckets in HwLD. As a by-product of our analysis, we also derive the limiting distribution of the maximum queue length in an $M|G|\infty$ queueing system. Our results generalize previous work in this area.

*Supported by N.S.F. grant MCS87-01426.

†Currently at the Department of Computer Science, University of Houston, Houston, TX 77204, U.S.A.

‡This research was supported by AFOSR grant 90-0107, in part by the N.S.F grant CCR-8900305, and in part by grant R01 LM05118 from the National Library of Medicine.

1. INTRODUCTION

The purpose of this paper is to present a thorough analysis of *Hashing with Lazy Deletion* (HwLD) in a general probabilistic framework. Items arrive at a hashing table and need to be stored for some period (the item's *life-time*). Different probability models for arrival and life-times are discussed later. We always assume that the assignment of items to the H buckets of the hashing table is uniform: that is, each item has probability $1/H$ to select each bucket, independent for different items and independent of the arrival and life-times.

The strategy of HwLD was proposed by Van Wyk and Vitter [22]. The principle of HwLD is very simple, namely: an item in a bucket is not deleted as soon as possible (i.e., when its life-time expires). Instead, the item is removed at the first arrival to the item's bucket following the item's expiration time. The point is that algorithms which delete items as soon as possible may have unacceptably high overhead, even though they require less storage space for the items themselves. In other words, there is a tradeoff between the time overhead incurred by immediate deletions and the space overhead that accrues if we want to keep the time overhead small. For more details concerning HwLD and its applications the reader is referred to [22, 16, 17, 18].

A natural problem to examine is how much storage space HwLD requires, and compare it with the storage space of a standard hashing strategy that we shall call *Hashing with Immediate Deletion* (HwID). A particularly intriguing problem is to estimate the amount of excess space used by HwLD. Let $U_H(t)$ and $N_H(t)$ denote the number of items at time t in a table with H buckets, used for HwLD and HwID respectively; think of this notation as a mnemonic for the 'used' and 'needed' amounts of space. The term "table size" will be conventionally used to denote either of these quantities. Let $W_H(t) \equiv U_H(t) - N_H(t)$ be the space that the HwLD wastes at time t. We investigate the (expected) instantaneous difference $E[W_H(t)]$, and the difference between $E\max_{0\le t\le T} U_H(t)$ and $E\max_{0\le t\le T} N_H(t)$. These two differences are called the (expected) "wasted space" and "excess space", respectively. Also, there is interest in evaluating $\max_{0\le t\le T} N_H(t)$ and $\max_{0\le t\le T} U_H(t)$ themselves. To motivate this further we note – after Van Wyk and Vitter [22] – that $N_H(t)$ can be interpreted as the number of "live" items at time t, regardless of the hashing strategy implementation. In other words, $N_H(t)$ is the minimum space requirement for any algorithm that maintains $N_H(t)$ items in the data structure at time t. For such problems the quantity $\max_{0\le t\le T} N_H(t)$ is a lower bound on the space requirement, and $\max_{0\le t\le T} U_H(t)$ is the corresponding space used by hashing with lazy deletion. We shall show that both display similar growth with respect to the traffic intensity and time. Furthermore, the difference, $\max_{0\le t\le T} U_H(t) - \max_{0\le t\le T} N_H(t)$ will be shown to be small, in a sense we detail below: hence the HwLD strategy can be said to be *near optimal* in terms of storage-space requirements [22], and very attractive from the time complexity viewpoint due to its low overhead cost. We study these and some related questions in this paper.

Although this paper adopts a queueing-theoretical approach, it differs from the traditional queueing analyses in some important aspects. Our look at the problem resembles the one studied by Morrison, Shepp and Van Wyk [16]; that is, we first consider a model suitable for a single bucket, and then we analyze the complete model, involving a (finite) number of such buckets. We use a natural sample-path approach that gives readily answers concerning the average wasted space problem in HwLD. To study the *excess* space we have to evaluate the maximum queue length in $GI|G|\infty$ queueing systems[1], and we prove some new results concerning this maximum. In passing we note that while we only consider hashing tables, the evaluation of maximum queue-lengths might be useful for the analysis of several other data structures. Our methodology can be applied to study dynamics of data structures that share some common features with *queues*, namely structures that are built during a sequence of insertions and deletions [9, 15, 14]. We mention here dictionaries, linear lists, stacks, priority queues and symbol tables [3].

The literature on hashing with lazy deletion is rather scanty. As mentioned above, HwLD was introduced by Van Wyk and Vitter [22]. Under exponential/exponential interarrival/lifetimes assumptions ($M|M|\infty$ model) they proved that $EU_H(t) - EN_H(t) = H$. For the same model, Morrison, Shepp and Van Wyk [16] estimated *numerically* the distribution of $\max_{0\le t\le T} U_H(t)$, and from these numerical analyses they conjectured that the difference $E\{\max_{0\le t\le T} U_H(t)\} - E\{\max_{0\le t\le T} N_H(t)\} = O(H)$. In two recent papers Mathieu and Vitter [17,18] proved this conjecture for an $M|G|\infty$ model, using an interesting (and rather complicated) probabilistic approach. In addition, [18] establishes the rate of growth for the maximum queue length in $M|G|\infty$ model. Some preliminary results concerning HwLD are also presented in Szpankowski [21]. Our results provide generalizations in various directions. First, we investigate the most general $GI|G|\infty$ model, and obtain basic results in this setting. In particular, we show how they differ from the $M|G|\infty$ model. We prove—as conjectured—that indeed $EU_H(t) - EN_H(t) = H$ in

[1] A typical single queueing model is that of $GI|G|c$ where the first G stands for general (arbitrary) interarrival time distribution of items (customers), the second G denotes the general (arbitrary) life time distribution, and finally c represents the number of servers. When an I is affixed to the first G it signifies that the interarrival duration distribution is sampled independently each time. Finally, $M|G|\infty$ denotes the specialization in which the arrival time process is Poisson with rate λ, and $GI|M|\infty$ denotes the specialization in which the life-time distribution is exponential (μ), and with an infinite number of servers [13].

the $M|G|\infty$ model of HwLD (see also [18]) but *not* in the $GI|M|\infty$ model (Theorem 1). Next we consider the maximum table size under HwLD, and prove that in general $\max_{1\le k\le n} U_H(\tau_k-) = o(\log n)$, where τ_k is the arrival time for the k-th item, and in particular for $M|G|\infty$ (see also [18]) that $\max_{1\le k\le n} U_H(\tau_k-) \sim \log n/\log\log n$ (Theorem 2). Finally, we deal with the excess space and prove that in the $M|G|\infty$ model of HwLD, $\Pr\{\max_{1\le k\le n} U_H(\tau_k-) - \max_{1\le k\le n} N_H(\tau_k-) > H+2\} \to 0$ as $n\to\infty$ (Theorem 3(i)). To derive this result we need to obtain sharp asymptotics for the distribution of the maximum queue length in an $M|G|\infty$ queue (Theorem 6) which seems to be a new result. We have also one result on the excess space for the general model without any probability assumptions on arrival and life-times, namely: for large H, and n *polynomially* large in H, we show that $\Pr\{\max_{0\le k\le n} U_H(\tau_k-) - \max_{0\le k\le n} N_H(\tau_k-) \ge H + O(\sqrt{H\log n})\} = o(1)$ for large n (Theorem 3(ii)).

The paper is organized as follows. In the next section we formulate a probabilistic model of HwLD, and state our results. Section 3 contains the proofs of those results that deal with the maximum size. These proofs require us to investigate the asymptotic distribution of the maximum queue length in a queueing system with an infinite number of servers. Finally, in section 4 we sketch future research directions that aim to get a more realistic approach to the maximum size of dynamic data structures.

2. STATEMENT OF RESULTS

We consider a table managed under HwLD with H buckets. Items arrive at arbitrary times $0\le\tau_1<\tau_2<\dots$. Let $\eta(t)$ represent the number of arrivals up to time t. An arriving item selects one out of the H buckets at random (with uniform probability) and joins the items assigned to this bucket. The k'th item has a life-time (required storage time) $S_k>0$. Under Hashing with Immediate Deletion (HwID), the k'th item is removed at time τ_k+S_k. Let $N_H(t)$ be the total number of items in the hash table at time t under HwID, and let $N_H^{(i)}(t)$ be the number in bucket i. Under the Hashing with Lazy Deletion scheme with the same arrival and life-times, let $U_H(t)$ be the total number of items in the hash table at time t, and let $U_H^{(i)}(t)$ be the number in bucket i. Let $W_H(t)=U_H(t)-N_H(t)$ denote the wasted space.

¿From the verbal description of HwLD and Figure 1, we see the following *sample path* relationship in each bucket i and at each time t

$$U_H^{(i)}(t) = N_H^{(i)}(\tau^{(i)}_{\eta(t)}-)+1 \tag{2.1}$$

where $\tau^{(i)}_{\eta(t)}$ denotes the time of the last arrival to bucket i before time t. Note that strictly

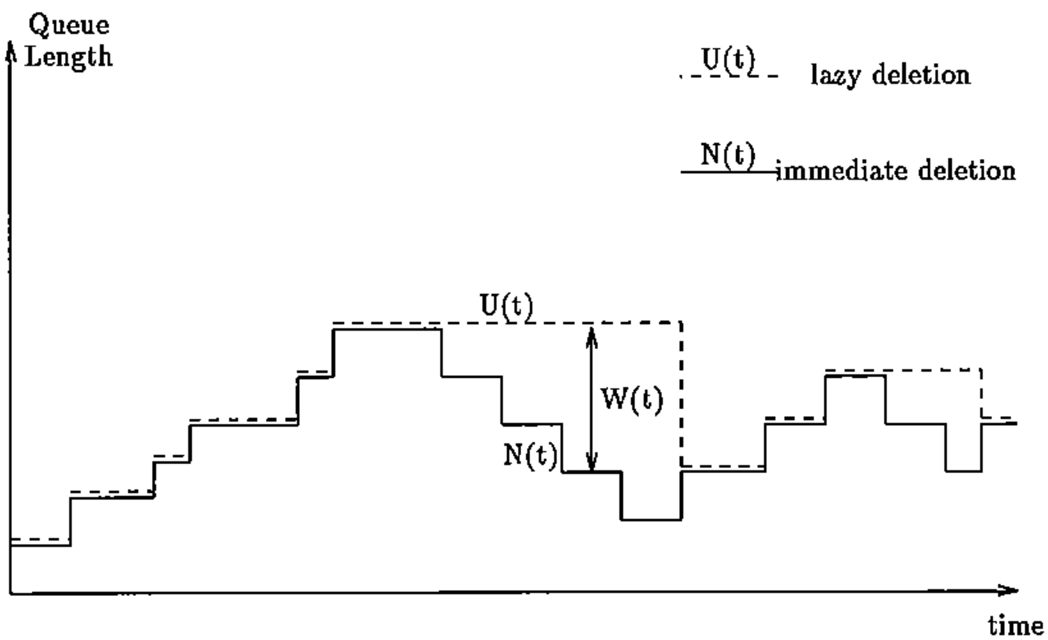

Figure 1: Relationship between $N_H(t)$ and $U_H(t)$ in a single bucket

speaking, (2.1) only holds after the first arrival of a customer to the queue. So $N_H^{(i)}(\tau_{\eta(t)}^{(i)}-)$ denotes the number of items in bucket i with unexpired life-times, immediately before the time of the last arrival to bucket i before t (i.e. the number *seen* by that arrival). Summing over buckets,

$$U_H(t) = \sum_{i=1}^{H} N_H^{(i)}(\tau_{\eta(t)}^{(i)}-) + H, \tag{2.2}$$

Similar to the restriction on equation (2.1), also (2.2) only holds after every bucket has had an arrival. Thus (2.2) expresses the number $U_H(t)$ of items used by HwLD in terms of the queue-length processes $N_H^{(i)}(t)$ in individual buckets with immediate deletion.

So far we have made no assumptions about the arrival and life-times, and in this generality we have only one result (Theorem 3(ii)). For the other results we introduce probabilistic models for the arrival and life-times. In the $GI|G|\infty$ model, the interarrival times $\xi_k = \tau_k - \tau_{k-1}$ are assumed to be strictly positive independent and identically-distributed (i.i.d.) random variables with mean $1/\lambda$, and the life-times S_k are also assumed to be strictly positive i.i.d. with mean $1/\mu$. Let $\rho = \lambda/\mu$ denote the traffic intensity.

We state our results for the *stationary* versions of these processes. An alternative is to assume the table starts empty. The results about asymptotic maxima (Theorem 2 and Theorem 3(i)) are unchanged, whereas Theorem 1 would hold with N_H and U_H interpreted as the limit ($t \to \infty$) in distribution of $N_H(t)$ and $U_H(t)$. Note that this limit exists under weak technical assumptions using regeneration arguments [11].

It is important to note that the processes $N^{(i)}(t)$ are in general *dependent* as the bucket i varies (and similarly for $U^{(i)}(t)$). The $M|G|\infty$ model is an exception: by the "independent sampling" property of the Poisson arrival process, what happens in different buckets is *independent.*

Now we are ready to present our results concerning hashing with lazy deletion. We concentrate on comparing it with ordinary hashing, that is, with immediate deletion.

THEOREM 1. Stationary Distribution and Moments of the Table Content. *Consider the stationary $GI|G|\infty$ model of HwLD. Let U_H and N_H be the limiting random variables for $U_H(t)$ and $N_H(t)$. Let $N^{(i)}(\tau^{(i)}-)$ denote the number of items seen in bucket i by an item arriving in bucket i, in the corresponding* immediate deletion *model.*

(i) In the $GI|G|\infty$ model

$$\Pr\{U_H = k + H\} = \Pr\{\sum_{i=1}^{H} N^{(i)}(\tau^{(i)}-) = k\}; \qquad k \geq 0 \tag{2.3a}$$

and

$$EU_H = H(1 + EN^{(i)}(\tau^{(i)}-)), \quad \textit{for any fixed } i\ . \tag{2.3b}$$

(ii) In the $M|G|\infty$ model $U_H - H$ and N_H each have Poisson (ρ) distribution, that is,

$$\Pr\{U_H = k + H\} = e^{-\rho}\frac{\rho^k}{k!}\ , \qquad k \geq 0. \tag{2.3c}$$

So in particular

$$EU_H = EN_H + H \tag{2.3d}$$

$$\mathrm{var}U_H = \mathrm{var}N_H\ . \tag{2.3e}$$

(iii) In the $GI|M|\infty$ model

$$EU_H = \frac{1}{\rho}\frac{A^*(\mu)}{1 - A^*(\mu)}EN_H + H\ , \tag{2.3f}$$

where $A^*(\mu) = Ee^{-\mu\xi}$ and ξ is the inter-arrival time.

Remark. Note that (2.3f) implies that (2.3d) does *not* in general hold for non-Poisson arrival processes.

Proof. Part (i) is immediate from (2.2). For (ii), N_H has the stationary distribution of the $M|G|\infty$ queue, which is well known to be Poisson (ρ). Applying (2.1) to bucket i, and using the PASTA[2] property, we see that $U^{(i)} - 1$ has Poisson (ρ/H) distribution. Summing over buckets (and using independence between buckets) gives (2.3c), which immediately implies (2.3d,e). This was also obtained in [18], by a rather different and more complicated manner.

To prove (iii), note that in the $GI|M|\infty$ queue $EN_H = \rho$ [23, p. 348]. So in view of (2.3b), what we need to show is

$$EN^{(i)}(\tau^{(i)}-) = \frac{A^*(\mu)/H}{1 - A^*(\mu)} .$$

Now the immediate-deletion process in the given bucket i is the $GI|M|\infty$ queue with a different inter-arrival time $\hat{\xi}$, say. Let $\hat{A}^*(u) = Ee^{-u\hat{\xi}}$. A standard computation [23, ibid.] (conditioning on all previous arrival times) gives the probability generating function (pgf) of $N^{(i)}(\tau^{(i)}-)$

$$Ez^{N^{(i)}(\tau^{(i)}-)} = \exp\{\frac{\hat{A}^*(\mu)}{1 - \hat{A}^*(\mu)} \log z\} .$$

In particular,

$$EN^{(i)}(\tau^{(i)}-) = \frac{\hat{A}^*(\mu)}{1 - \hat{A}^*(\mu)} .$$

Now $\hat{\xi}$ is a sum of a geometrically-distributed number of inter-arrival times ξ_j

$$\hat{\xi} = \sum_{j=1}^{G} \xi_j \; ; \qquad P(G = g) = H^{-1}(1 - 1/H)^{g-1}, \quad g \geq 1$$

and a brief calculation gives

$$Ee^{-\mu\hat{\xi}} = \frac{A^*(\mu)/H}{1 - A^*(\mu)(1 - 1/H)}.$$

Substituting this into the previous formula leads to the desired equation, completing the proof of Theorem 1. ■

Our main results concern the maximum table size over long time intervals. Note that the time of attainment of the maximum (for either $N_H(t)$ or $U_H(t)$) must occur immediately after some arrival. Thus we can state Theorems 2 and 3 in terms of maxima seen at arrival times, and the results remain true also if we interpret the maxima as taken over the entire corresponding time intervals—up to a difference of one, since in the latter case the arrival is counted as well.

[2]PASTA stands for *Poisson Arrivals Sees Time Average*, and this implies that the time-stationary distribution of the queue length is the same as the customer-stationary distribution, that is, as seen by an arriving customer. More details can be found in [23], mainly section 5.16.

The first result of this type defines the order of growth with time of the maximum occupancy of the table using HwLD. The proof is given in section 3. To review some standard notation, $a_n \sim b_n$ means $a_n/b_n \to 1$, and for random variables $X_n \to 0$ *in probability* (pr.) means $\Pr\{|X_n| > \varepsilon\} \to 0$ as $n \to \infty$ for any fixed $\varepsilon > 0$. The symbol $\lfloor x \rfloor$ denotes the largest integer smaller than or equal to x.

THEOREM 2. Maximum Size of a Table under HwLD.

(i) For an $M|G|\infty$ model of HwLD, suppose the life-time S satisfies $ES\log^2 S < \infty$. Then,

$$\Pr\{\lfloor a_n \rfloor + 1 \le \max_{1\le k\le n} U_H(\tau_k) \le \lfloor a_n \rfloor + 1 + H\} \to 1 \ \ as \ \ n \to \infty \ , \tag{2.4a}$$

where $\{a_n\}$ is a particular sequence defined below, which satisfies $a_n \sim \log n/\log\log n$.

(ii) For a $GI|G|\infty$ model of HwLD the maximum table size satisfies $\max_{1\le k\le n} U_H(\tau_k) = o(\log n)$ in probability; precisely, as $n \to \infty$

$$\frac{1}{\log n}\max_{1\le k\le n} U_H(\tau_k) \to 0 \quad (pr.) \ , \tag{2.4b}$$

provided the life-time S satisfies $\Pr\{S > x\} = O(e^{-\beta x})$ for some $\beta > 0$. ■

Remark. Part (i) implies that $\max_{1\le k\le n} U_H(\tau_k) \sim \log n/\log\log n$ (pr.) (also obtained in [18]). It is plausible that the conclusion $\max_{1\le k\le n} U_H(\tau_k) \sim c\log n/\log\log n$ (pr.), for some $c > 0$, also holds for the $GI|G|\infty$ model, under weak assumptions on inter-arrival and service times.

The next finding is our strongest result, and it estimates the *excess space* that HwLD requires in order to accommodate the same arrival process as ordinary hashing with immediate deletion. This result resolves some open problems posed in [22] and [16]. It also says that under fairly general assumptions HwLD is *near optimal.* Indeed, we prove the following.

THEOREM 3. Limiting Excess Space.

(i) In the stationary $M|G|\infty$ model of HwLD, as $n \to \infty$

$$\Pr\{\max_{1\le k\le n} U_H(\tau_k) - \max_{1\le k\le n} N_H(\tau_k) > H + 2\} \to 0 \ , \tag{2.5a}$$

provided the life-time S satisfies $ES\log^2 S < \infty$.

(ii) Consider HwLD with arbitrary (i.e., no probabilistic assumptions) arrival and life-times. Then for $n, H \ge 2$ and $b > H$,

$$\Pr\{\max_{1\le k\le n} U_H(\tau_k) - \max_{1\le k\le n} N_H(\tau_k) \ge b\} \le 2n\left(\frac{H+b}{2b}\right)^{b/2}\left(\frac{H+b}{2H}\right)^{H/2} \ . \tag{2.5b}$$

If H is large and n is at most polynomially large in H, then the bound on the difference is $H + O(\sqrt{H \log n})$. In particular, $\Pr\{\max_{0 \le k \le n} U_H(\tau_k) - \max_{0 \le k \le n} N_H(\tau_k) > H + (2 + \varepsilon)\sqrt{H \log n}\} = o(1)$, *for any* $\varepsilon > 0$. ∎

In summary, our results indicate that hashing with lazy deletion should provide a very attractive alternative solution for hashing implementations. In particular, under fairly general probabilistic assumptions the average storage space required by HwLD is *not* much larger than for an ordinary hashing with immediate deletion (Theorem 1). We would assume this observation would hold for a wider range of probabilistic models than those for which we could manufacture a proof. Furthermore, with very high probability, the excess space incurred by lazy deletion is relatively small compared with the space requirements of HwID (Theorem 3). While it increases with the life-time of the system, the rate of growth, $O(\sqrt{\log n})$, is reassuringly moderate. Since HwLD allows us to use data structures that have low space overhead, we are led to the conclusion that hashing with lazy deletion is *essentially optimal* in terms of space and time complexity. Note, however, that with small probability something may still go wrong with HwLD. Indeed, it is not difficult to create realizations in which the arrival and life-time processes interact to have time points at which the wasted space, i.e. the difference $U_H(t) - N_H(t)$, assumes arbitrarily large values.

Finally, one usually interprets $n \to \infty$ asymptotics as approximations for large finite n. The results we report here need sometimes a more precise statement about the relation between the parameters. For example, some results would require n to be "super-exponentially large in ρ" in order for the approximation to be valid. In such a case the asymptotic results have limited practical importance. We shall comment on this difficulty, and suggest an alternative approach for its resolution, in our concluding remarks in Section 4.

3. ANALYSIS OF THE MAXIMUM SIZE

In this section we prove Theorems 2 and 3 stated above. Both theorems deal with the maximum size of a table under HwLD. In the course of deriving these results we present some new findings concerning an asymptotic distribution of the maximum queue length in an $M|G|\infty$ queue (Theorem 6).

3.1 Maximum Size of HwLD

To obtain the required bounds on the table size under HwLD, the following Lemma, Corollary and Theorem show progressively tighter bounds on the maxima of sequences of identically distributed random variables. Lemma 4 (and its Corollary 5) are a direct consequence of Anderson's findings [5], but we bring them here for convenience of reference.

Lemma 4. *Let X_1, X_2 ,... be identically distributed* discrete, possibly dependent *random variables with common marginal distribution function $F(x) = \Pr\{X < x\}$ where x belongs to the set $\mathcal{N}$ of nonnegative integers. We denote $M_n \equiv \max_{1\le k\le n} X_k$.*

(i) Let

$$F(x) < 1 \qquad \text{for} \quad x < \infty \ , \tag{3.1}$$

and assume a function $g(x,b)$ exists, such that for any positive integer $b \in \mathcal{N}^+$ the distribution function $F(x)$ satisfies

$$\frac{1 - F(b + x)}{1 - F(x)} = g(x,b) \tag{3.2}$$

where $\lim_{x\to\infty} g(x,b) = 0$, (that is, the distribution of X_i has an superexponential tail). Let also a_n be the smallest solution of the characteristic equation [3]

$$n[1 - F(a_n)] = 1 \ . \tag{3.3}$$

Then,

$$\Pr\{M_n \ge \lfloor a_n \rfloor + 1 + b\} = O(g(a_n, b)) \to 0 \ , \quad n \to \infty \tag{3.4}$$

In other words, $M_n \le \lfloor a_n \rfloor + 1$ (pr.)

(ii) If X_1, X_2 ,..., X_n are **independent** *random variables satisfying the above hypotheses, then*

$$\Pr\{M_n < x\} - \exp(-n[1 - F(x)]) \to 0 \qquad as \quad n, x \to \infty \tag{3.5a}$$

and

$$\Pr\{M_n = \lfloor a_n \rfloor + 1 \ or \ \lfloor a_n \rfloor\} = 1 - O(g(a_n, 1)) \to 1 \qquad as \ \ n \to \infty \ , \tag{3.5b}$$

where a_n solves (3.3).

Proof. (i) Equation (3.4) follows directly from Boole's inequality and the superexponentiality assumption (3.2), namely for $b \in \mathcal{N}^+$

$$\Pr\{M_n \ge \lfloor a_n \rfloor + b + 1\} \le n \cdot [1 - F(\lfloor a_n \rfloor + b + 1)] = O(g(a_n, b)) \ \to \ 0$$

[3] Since the distribution function is only piece-wise continuous, with jumps at the integers, equation (3.3) may not be satisfiable for any n. We define then a "solution" of (3.3) by embedding the discrete random variables in a continuous version with a distribution that coincides with $F(x)$ at the integers. Following [5], let $G(x) = 1 - F(x)$, $h(n) = -logG(n)$ and $h_c(x) \equiv h(\lfloor x \rfloor) + (x - \lfloor x \rfloor)(h(\lfloor x \rfloor + 1) - h(\lfloor x \rfloor))$. Then the continuous complementary distribution $G_c(x) \equiv e^{-h_c(x)}$ is the function we use; a_n is the solution of $G_c(a_n) = 1/n$.

when $a_n \to \infty$, which follows from (3.1) and (3.2). This sequence $\{a_n\}$ is the one used in the formulation of Theorem 2.

(ii) Let $G(x) = 1 - F(x)$. Equation (3.5a) follows immediately from the observation that $\Pr\{M_n < x\} = F^n(x)$, and developing it as

$$\Pr\{M_n < x\} - e^{-nG(x)} = e^{-nG(x)}(e^{-nG^2(x)(1/2+G(x)/3+\ldots)} - 1).$$

It can be seen that either of the two factors on the right-hand side vanishes as x or n increases.

For (3.5b) we note that since M_n assumes integer values only, we may write

$$\Pr\{M_n < \lfloor a_n \rfloor\} = \Pr\{M_n < \lfloor a_n \rfloor - \varepsilon\} \le \Pr\{M_n < a_n - \varepsilon\}$$

for some $0 < \varepsilon < 1$ whether a_n is integer or not. Then from relation (3.5a) we have for n large enough, where $G_c(x)$ is a continuous version of $G(x)$ (see last footnote),

$$\Pr\{M_n < \lfloor a_n \rfloor\} \le \exp\{-nG_c(a_n - \varepsilon)\} = \exp\{-\frac{G_c(a_n - \varepsilon)}{G_c(a_n)}\}.$$

Since for $G_c(x)$ the analogue of equation (3.2) holds for any $b > 0$, the last argument in braces is unbounded as $n \to \infty$, and hence

$$\Pr\{M_n < \lfloor a_n \rfloor\} \;=\; o(1).$$

This, together with part (i), imply the result. ■

As a direct consequence of the above we show the following corollary concerning the maximum of the Poisson process.

Corollary 5. *(i) Let $\{X_k; k \ge 1\}$ be (possibly dependent) Poisson(ρ) variables. Let $I(n)$ be a random sequence possibly dependent on the $\{X_k\}$, with $I(n)/n \to c$ (pr.) as $n \to \infty$, for some finite $c > 0$. Then there exists (an increasing) sequence x_n satisfying the following*

$$\Pr\{\max_{1 \le k \le I(n)} X_k < x_n\} \to 1 \;.$$

(ii) If $\{X_k, k \ge 1\}$ are i.i.d. Poisson (ρ) distributed random variables, then for large enough integers a and n

$$\Pr\{\max_{1 \le k \le n} X_k < a\} - \exp(-ne^{-\rho}\rho^a/a!) \to 0 \quad as \quad n, a \to \infty \;, \tag{3.6a}$$

and

$$\Pr\{\max_{1 \le k \le n} X_k = \lfloor a_n \rfloor + 1 \text{ or } \lfloor a_n \rfloor\} = 1 - O(1/a_n) \to 1 \qquad as\ n \to \infty \;. \tag{3.6b}$$

For large n the sequence $\{a_n\}$ satisfies

$$a_n \sim \frac{\log n - \rho}{\log(\log n - \rho) - \log \rho} \sim \frac{\log n}{\log \log n} , \tag{3.6c}$$

where a_n is defined as the smallest solution of the equation

$$n \cdot \frac{\gamma(a_n, \rho)}{\Gamma(a_n)} = 1 . \tag{3.7a}$$

In the above $\gamma(x,\rho) \equiv \int_0^\rho t^{x-1}e^{-t}dt$ is the incomplete gamma function, and $\Gamma(x) = \gamma(x,\infty)$ is the gamma function [1].

Remark. Using the property $P(X \geq x) \sim P(X = x)$ we can specify the sequence in (3.6b) in an alternative way. Namely, (3.6b) holds with $\lfloor a_n \rfloor$ replaced by any integer-valued sequence a_n satisfying

$$ne^{-\rho}\rho^{a_n+2}/(a_n+2)! \to 0 \qquad \text{and} \qquad ne^{-\rho}\rho^{a_n}/(a_n)! \to \infty , \tag{3.7b}$$

and $a_n \to \infty$.

Proof. Part (i) follows from the same arguments as in the proof of Theorem 3.2 in Berman [6] (see also [5, pp. 109-111], [10 Chap. 6.2]). In particular, using the partition arguments of [6] for any $\varepsilon > 0$ we obtain $\Pr\{\max_{1 \leq k \leq I(n)} X_i > x_n\} \leq 2\varepsilon + nc(1+\varepsilon)\Pr\{X_i > x_n\}$, as in the proof of Lemma 4(i). Putting $x_n = \lfloor a_{\lfloor nc \rfloor} \rfloor + 2$, with a_n given by (3.6c) we establish part (i).

For part (ii), equation (3.6a) is immediate from (3.5a), on observing that $\Pr\{X \geq x\} \sim \Pr\{X = x\}$, as $x \to \infty$ (due to superexponentiality of the Poisson distribution). Equation (3.6b) is identical with equation (3.5b), and for the value of a_n one needs only to notice that the tail of the Poisson distribution can be computed as

$$1 - F(x) = \Pr\{X \geq x\} = \sum_{j=x}^{\infty} \frac{\rho^j}{j!} e^{-\rho} = \frac{\gamma(x,\rho)}{\Gamma(x)} .$$

For an asymptotic solution of (3.7a) we follow [1, p. 262] and approximate the incomplete gamma function as $\gamma(x,\rho) \approx \Gamma(x)e^{-\rho}\rho^x/\Gamma(x+1)$. Hence for large n equation (3.7a) reduces to

$$n \cdot \frac{e^{-\rho}\rho^{a_n}}{\Gamma(a_n+1)} = 1 . \tag{3.7c}$$

Applying Stirling's formula to the above, one finds

$$\left(\frac{a_n}{\rho e}\right)^{a_n} = \frac{e^{-\rho}}{\sqrt{2\pi}} \frac{n}{\sqrt{a_n}} .$$

This equation can be solved for large n by *asymptotic bootstrapping*, and this leads to equation (3.6c). Finally, the evaluation of the function $g(x,1)$ from Lemma 4 gives $g(x,1) = \rho x!/(x+1)! \sim \rho/x$ and this gives (3.6b). ∎

In order to prove Theorem 2(i) and Theorem 3(i) we need sharp asymptotic estimates for the maximum queue length in an $M|G|\infty$ queue. Recall that the queue length in $M|G|\infty$ has the stationary Poisson (ρ) distribution; however, the dependence of queue sizes at different times precludes the simple-minded use of Corollary 5. Note also that the queue-length process in *not* Markov. We shall prove the following theorem and show that it implies, together with Theorem 3(i), directly Theorem 2(i).

THEOREM 6. *Let X_t be the queue length in the stationary $M|G|\infty$ queue. Then, uniformly in t_0,*

$$\Pr\{\sup_{t \le t_0} X_t \le a\} - \exp(-t_0 \lambda e^{-\rho} \rho^a / a!) \to 0 \quad as \quad a \to \infty \tag{3.8a}$$

Now let $X_{\tau_k -}$ be the queue length X_t just before the arrival time τ_k. Then, with the sequence a_n that provides the solution to equation (3.7a) we find

$$\Pr\{\max_{1 \le k \le n} X_{\tau_k -} = \lfloor a_n \rfloor + 1 \ or \ \lfloor a_n \rfloor\} \to 1 \ as \ n \to \infty \ , \tag{3.8b}$$

provided the life-time S satisfies $ES \log^2 S < \infty$.

Remark. One could reformulate equation (3.8b) to refer to a maximum on a time interval T. This would lead to a similar right-hand side, with a_n replaced by $a_{\lfloor \lambda T \rfloor}$.

Proof. Consider first relation (3.8a). Fix an integer a. Call a time t with $X_{t-} = a$ and $X_t = a+1$ an *upcrossing time.* Classify items in the queue as "cleared" or "uncleared" according to the following rules.

(i) Each new arrival is "uncleared".
(ii) Whenever the number of "uncleared" items increases to $a+1$, all these $a+1$ items are declared "cleared". (Call such an time a *clearing time.*)

There is a stationary version of this process, and for this stationary version define
X_t = total number of items at time t
X_t^* = number of uncleared items at time t.
Of course X_t by itself is the $M|G|\infty$ queue. And (X_t^*) by itself can be regarded as the process which behaves like the $M|G|\infty$ queue with the following modification: when an arrival makes the queue length equal to $a+1$, all items in storage are removed. The purpose of the joint construction is to obtain the following property.

(iii) the set of clearing times for X_t^* is a subset of the set of upcrossing times for X_t.

To see why, let t_1 be a clearing time and let t_0 be the last time before t_1 that the queue was empty. Then $X_t = X_t^*$ on $t_0 \le t < t_1$, and so t_1 is an upcrossing time for X_t.

Write $q(a)$ for the chance that a typical upcrossing time of X is a clearing time of X^*. Then

$$q(a) = \frac{\text{rate of clearings of } X_t^*}{\text{rate of upcrossings of } X_t} = \frac{1}{E_0 T_{a+1}} \times \frac{a!}{\lambda e^{-\rho}\rho^a}$$

where T_{a+1} denotes the first hitting time

$$T_{a+1} = \min\{t : X_t = a+1\}$$

for the $M|G|\infty$ process, and E_0 (and later Pr_0) indicate quantities that refer to a process started at state 0 (i.e. empty). By (iii), $q(a) \le 1$. The key fact, proved in the Appendix by a different argument, is the following lemma.

Lemma 7. *Provided $ES \log^2 S < \infty$, we have $q(a) \to 1$ as $a \to \infty$.* ■

Because the $M|G|\infty$ queue regenerates at state 0, a standard argument [12] gives an exponential limit distribution for hitting times:

$$\mathrm{Pr}_0\{T_{a+1} > sE_0 T_{a+1}\} \to e^{-s} \qquad \text{as } a \to \infty, \text{ uniformly in } s.$$

This implies that the point process of clearing times of X^*, with time normalized by $E_0 T_{a+1}$, converges (as $a \to \infty$) to a Poisson point process of rate 1. Lemma 7 now implies that the point process of upcrossings of the *stationary* queue X_t undergoes the same convergence. In particular, the (rescaled) time of the first upcrossing of X_t converges in distribution to the time of the first event of the Poisson process:

$$\Pr\{T_{a+1} > sE_0 T_{a+1}\} \to e^{-s} \qquad \text{as } a \to \infty, \text{ uniformly in } s \tag{3.9}$$

(which differs from the previous assertion, because it concerns the queue started with the stationary queue-size distribution, rather than a queue started empty). The uniformity in (3.9) and below is a consequence of the elementary fact that, in the context of convergence of distribution functions to a continuous distribution function, pointwise convergence implies uniform convergence.

Defining $s = s(a, t_0)$ by

$$s \frac{a!}{\lambda e^{-\rho}\rho^a} = t_0,$$

we can restate (3.9) as

$$\Pr\{T_{a+1} > sE_0 T_{a+1}\} - \exp(-t_0 \lambda e^{-\rho}\rho^a / a!) \to 0 \text{ as } a \to \infty, \text{ uniformly in } t_0.$$

Now $sE_0T_{a+1} \to t_0$ by Lemma 7, so

$$\Pr\{T_{a+1} > t_0\} - \exp(-t_0\lambda e^{-\rho}\rho^a/a!) \to 0 \text{ as } a \to \infty, \text{ uniformly in } t_0.$$

But this gives (3.8a), since the events $\{T_{a+1} > t_0\}$ and $\{\max_{t\leq t_0} X_t \leq a\}$ are the same, provided $X_0 \leq a$.

Equation (3.8b) is an immediate result of (3.8a) and the definition of a_n (cf. (3.7)), since it provides for $\Pr\{M_n > \lfloor a_n \rfloor - 1\} \to 1$ and $\Pr\{M_n < \lfloor a_n \rfloor + 2\} \to 1$. ■

The rest of this subsection is devoted to the proof of Theorem 2(ii) regarding the size of HwLD under the $GI|G|\infty$ model. Our result will follow easily from the following estimate of the tail of the queue length in a $GI|G|\infty$ queue.

Lemma 8. *In the stationary $GI|G|\infty$ queue, let N be the number of customers seen by an arriving customer. Then*

$$\Pr\{N \geq n\} = o(\alpha^n) \text{ as } n \to \infty, \text{ for every } \alpha > 0$$

provided the life-time S satisfies $\Pr\{S > x\} = O(e^{-\beta x})$ *for some* $\beta > 0$.

Proof. Consider the stationary queue, conditioned on an arrival at time $\tau_0 = 0$. The previous arrivals were at times $(-\tau_1, -\tau_2, \ldots)$, where $\tau_n = \sum_{i=1}^n \xi_i$, and the ξ_i are the inter-arrival times. Write $G(x) = \Pr\{S \geq x\}$, where S is the life-time. The distribution of N, the number of customers seen by the arriving customer at time 0, when the previous arrival times are given, can be described as

$$\text{for } \textit{given } (\tau_1, \tau_2, \ldots),\ N \text{ is distributed as } \sum_{i=1}^{\infty} 1_{A_i},$$
$$\text{where the } A_i \text{ are } \textit{independent} \text{ and } \Pr\{A_i\} = G(\tau_i)\ .$$

(Here A_i is the event that the customer who arrived at time $-\tau_i$ is still present at time 0.)

Fix now $t_0 > 0$. Split N as $N_1' + N_2'$, where N_1' is the part of the sum $\sum 1_{A_i}$ over those i with $\tau_i \leq t_0$, and where N_2' is the part of the sum $\sum 1_{A_i}$ over those i with $\tau_i > t_0$. Obviously

$$N_1' \leq N_1 \equiv \max\{n : \tau_n \leq t_0\}\ ,$$

and N_2' has distribution described by

$$\text{for } \textit{given } (\tau_1, \tau_2, \ldots),\ N_2' \text{ is distributed as } \sum_{i=N_1+1}^{\infty} 1_{A_i},$$
$$\text{where the } A_i \text{ are independent with } \Pr\{A_i\} = G(\tau_i).$$

Now the process $(\tau_{N_1+i}-t_0; i \geq 1)$ is just a delayed version of the renewal process $(\tau_i; i \geq 0)$. (*Delayed* means there is not necessarily an arrival at time 0.) Using the natural coupling between this delayed process and the un-delayed renewal process, and the fact that $G(\cdot)$ is decreasing, we can represent $N_2' \leq N_2$, where N_2 has distribution described by

$$\text{for given } (\tau_0, \tau_1, \tau_2, \ldots),\ N_2 \text{ is distributed as } \sum_{i=0}^{\infty} 1_{A_i}$$
$$\text{where the } A_i \text{ are independent with } \Pr\{A_i\} = G(t_0 + \tau_i).$$

Thus $N \leq N_1 + N_2$, and we analyze these terms separately.

We first show

$$\Pr\{N_1 \geq n\} = o(\alpha^n) \text{ as } n \to \infty, \text{ for every } \alpha > 0\ . \tag{3.10a}$$

Indeed, given α consider K sufficiently large that $\Pr\{\xi < t_0/K\} \leq \alpha/2$. Such a K exists, since $\Pr\{\xi_i > 0\} = 1$. Then, as $n \to \infty$

$$\Pr\{N_1 \geq n\} = \Pr\{\sum_1^n \xi_i \leq t_0\} \leq \binom{n}{K} \Pr^{n-K}\{\xi \leq t_0/K\} = o(2\Pr\{\xi \leq t_0/K\})^n\ ,$$

where the inequality above is a simple consequence of the fact that at most K of the ξ's can exceed t_0/K. This implies assertion (3.10a).

Now consider N_2. A standard method (see e.g. the discussion of large deviations in [7] section 1.9) of obtaining exponentially small tail bounds on a r.v. is by studying the moment generating function. In particular, we can use the general inequality $\Pr\{X \geq a\} \leq E[g(X)]/g(a)$ which holds for any nondecreasing function $g(\cdot)$. Set $g(x) = e^{\phi x}$, then $Eg(X)$ is the moment generating function of X. The idea of the following proof is to show that $Eg(N_2) = O(1)$, and then $\Pr\{N_2 > n\} = O(e^{-\phi(t_0)n})$ for some $\phi(t_0) \to \infty$ as $t_0 \to \infty$.

By hypothesis about the life-time distribution there exist $A < \infty$ and $\beta > 0$ such that

$$G(x) \leq Ae^{-\beta x} \qquad \text{for all } x.$$

Choose $\gamma < \infty$ such that

$$\gamma E e^{-\beta\xi} < \gamma - 1\ . \tag{3.10b}$$

Lemma 9. *For all sufficiently small $\theta > 0$,*

$$E\exp(\theta \sum_{i=0}^{\infty} e^{-\beta\tau_i}) \leq \exp(\theta\gamma)\ .$$

Proof. See the Appendix. ■

Consider now $\phi > 0$. Then,

$$\begin{aligned} E\exp(\phi N_2|\tau_0,\tau_1,\ldots) &= \prod_{i=0}^{\infty}(1+(e^{\phi}-1)G(t_0+\tau_i)) \le \prod_{i=0}^{\infty}(1+(e^{\phi}-1)Ae^{-\beta t_0}e^{-\beta\tau_i}) \\ &\le \exp((e^{\phi}-1)Ae^{-\beta t_0}\sum_{i=0}^{\infty}e^{-\beta\tau_i}) \qquad (3.10c) \end{aligned}$$

Now it is straightforward to find a function $\phi(t_0) \to \infty$ as $t_0 \to \infty$, and such that also

$$\theta(t_0) \equiv (e^{\phi(t_0)}-1)Ae^{-\beta t_0} \to 0 \ .$$

Taking expectations over the arrival times in (3.10c) and applying Lemma 9,

$$E\exp(\phi(t_0)N_2) \le \exp(\theta(t_0)\gamma) \ .$$

We have from the moment generating function approach, as $n \to \infty$

$$\Pr\{N_2 \ge n\} = O(\exp(-\phi(t_0)n)) \ .$$

Recall $N \le N_1 + N_2$. Putting $\alpha = \exp(-\phi(t_0))$ in (3.10a),

$$\Pr\{N \ge 2n\} \le \Pr\{N_1 \ge n\} + \Pr\{N_2 \ge n\} = O(\exp(-\phi(t_0)n)) \ ,$$

as $n \to \infty$. This establishes Lemma 8, because $\phi(t_0) \to \infty$. ∎

Returning to the proof of Theorem 2(ii), consider $U_H(t)$, where the number H of buckets is a fixed constant. By (2.2), for any $1 \le i \le H$ with $k \le n$,

$$U_H(\tau_k) \le H \cdot \max_{1\le j\le n} N_H^{(i)}(\tau_j^{(i)}) \ .$$

Hence $\Pr\{U_H(\tau_k) > n\} = o(\alpha^n)$ as well. It follows easily that $\Pr\{\max_{1\le k\le n} U_H(\tau_k) > a_n\} = n \cdot o(\alpha^{a_n})$, for any $\alpha > 0$. Pick $a_n = -\log_\alpha n$, for arbitrary $0 < \alpha < 1$, to find $\Pr\{\max_{1\le k\le n} U_H(\tau_k) > -\log_\alpha n\} = n \cdot o(1/n) = o(1)$. This proves (2.4b) since α can be arbitrary small.

3.3 Limiting Excess Space.

We now turn our attention to the evaluation of the excess space. We first prove Theorem 3(i) for a stationary $M|G|\infty$ model of HwLD, and then Theorem 3(ii) for arbitrary arrivals and life-times.

For an $M|G|\infty$ model of HwLD, let $U_H(\tau_k-)$ denote the table size just before the k'th arrival. By PASTA, we see that $U_H(\tau_k-) - H$ is distributed as $U_H(0) - H$, which by Theorem 1(ii) has the Poisson(ρ) distribution, for each k. Applying Corollary 5(i)

$$\Pr\{\max_{1\le k\le n} U_H(\tau_k-) \le a_n + 1 + H\} \to 1 \ , \qquad (3.11)$$

where a_n satisfies equations (3.7). Now $N_H(t)$ is the $M|G|\infty$ queue length process, and Theorem 6 provides sharp asymptotics for the maximum queue length in such a queue. Comparing (3.11) with (3.8b) one immediately obtains (2.5a) of Theorem 3(i).

Finally, we leave the realm of queueing models to prove Theorem 3(ii), which concerns the case of arbitrary deterministic arrival and departure times. First imagine the hashing table is empty at time 0. There are arrivals at arbitrary times $0 < \tau_1 < \tau_2 \ldots < \tau_n$ with departures at arbitrary times $\eta_k > \tau_k$. Fix n. The process $N_H(t)$ and the maximum $N^* \equiv \max_{k \leq n} N_H(\tau_k)$ are deterministic. The only *probabilistic element* is the choice of bucket on arrival. We first argue that the general case can be reduced to a certain special case. Regard the arrival times and assignments to buckets as fixed, but make the following modifications. First, put N^* items in the table at time 0, but make them all depart before τ_1. Then repeat the following procedure.

> If there is some departure at some time $\eta < \tau_n$ which causes $N_H(\eta) = N^* - 2$, then choose the first such η and delay the departure until a time immediately after the first arrival $\tau_j > \eta$ at which $N_H(\tau_j) = N^*$. (If there is no such time τ_j, then the item stays forever.)

It is easy to show that after a finite number of such changes, there will be no such departure time η. We are then in the *special case* where $N_H(\tau_1-) = N^* - 1$ and where arrivals and departures alternate, so that $N_H(t)$ alternates between $N^* - 1$ and N^* up until time τ_n. The point is that delaying an item's departure cannot decrease any $U_H(t)$. Theorem 3(ii) concerns an upper bound for $\max_{k \leq n} U_H(\tau_k)$ in terms of N^*. Going from the general case to the special case can only increase the former quantity, and leaves N^* unchanged. So it suffices to consider the special case.

Fix a time t just after an arrival, and look backwards in time from t. Let X_i be the bucket which contained the i'th-from-last departing item (before t). Let Y_i be the bucket which contained the i'th-from-last arriving item (before t). Write $f(i) = j$ to mean that the i'th-from-last departure was the j'th-from-last arrival. Then $f(i) > i$, by the alternation property. Let B_t be the number of excess items at time t. Such an item is one which was (say) the i'th departure before t, but has not yet been removed by subsequent arrivals. This requires

$$X_i \text{ is different from all of } Y_1, \ldots, Y_i.$$

So B_t is exactly the number of i's for which this holds. The next lemma abstracts the structure of B_t. Theorem 3(ii) follows by applying this lemma to $B = U_H(\tau_k) - N^*$, summing over $k = 1, \ldots, n$ and appealing to Boole's inequality.

Lemma 10. *Let $f : \{1,2,3,\ldots\} \to \{1,2,3,\ldots\}$ be a $1-1$ function with $f(i) > i$. Let $(Y_i; i \geq 1)$ be independent, uniform on $\{1,\ldots,H\}$. Let $X_i = Y_{f(i)}$. Let A_i be the event*

$$X_i \text{ is different from all of } Y_1,\ldots,Y_i.$$

Let B be the counting r.v. $B = \sum_{i\geq 1} 1_{A_i}$. Then, for $b > H$,

$$\Pr\{B \geq b\} \leq 2\left(\frac{H+b}{2b}\right)^{b/2}\left(\frac{H+b}{2H}\right)^{H/2}.$$

Proof. The proof uses the following martingale-type bound, which we prove first. A good modern reference for martingales and σ-fields is [7] Chapter 4. The martingale M_n we use is the "multiplicative" analog of the usual "additive" martingale associated with a sequence of events, the latter appearing e.g. in [7] Theorem 4.4.10.

Lemma 11. *Let A_i be events adapted to increasing σ-fields $(\mathcal{F}_i)$, $i \geq 1$. Let $B = \sum_{i\geq 1} 1_{A_i}$. Then*

$$\Pr\{B \geq b\} \leq 2\sqrt{\inf_{z>1} z^{-b} ED_z}$$

where

$$D_z = \prod_{i\geq 1} E(z^{1_{A_i}}|\mathcal{F}_{i-1}) = \prod_{i\geq 1}(1+(z-1)\Pr\{A_i|\mathcal{F}_{i-1}\}).$$

where we have used the conditional version of the expansion $Ez^{1_A} = 1 + (z-1)\Pr\{A\}$.

Proof of Lemma 11. It is enough to prove the bound for fixed $z > 1$ such that $ED_z < \infty$. Write $M_0 = 1$,

$$M_n = z^{\sum_{i=1}^n 1_{A_i}} / \prod_{i=1}^n E(z^{1_{A_i}}|\mathcal{F}_{i-1}),\ n \geq 1.$$

Then $\{M_n\}$ is a positive martingale. By the martingale convergence theorem ([7] Corollary 4.2.11) M_n converges a.s. to some limit r.v. M_∞, and $EM_\infty \leq EM_0 = 1$. Plainly M_∞ "ought to be" equal to z^B/D_z, and this is verified by noting that the denominator in the definition of M_n is increasing in n and converges a.s. to the a.s. finite limit D_z. Thus we have proved

$$E(z^B/D_z) \leq 1\,. \tag{3.12}$$

So

$$\begin{aligned}
\Pr\{B \geq b\} &\leq \Pr\{D_z > a\} + \Pr\{B \geq b, D_z \leq a\} \qquad \text{for any } a > 0\\
&\leq ED_z/a + \Pr\{z^B/D_z \geq z^b/a\} \leq ED_z/a + a/z^b \qquad \text{by (3.12)}\\
&\leq 2\sqrt{z^{-b}ED_z}, \qquad \text{putting } a = \sqrt{z^b ED_z} \quad \blacksquare
\end{aligned}$$

We evaluate now the required infimum. Let $\mathcal{F}_i$ be the σ-field generated by $(Y_j;\ j \leq i+1)$ and $(X_j; j \leq i)$. Then

$$\Pr\{A_i|\mathcal{F}_{i-1}\} = V_i/H, \qquad i \geq 1$$

where V_i is the number of values *not* taken by $Y_1, \ldots, Y_i$. So for $z > 1$ the quantity D_z is

$$D_z = \prod_{i \geq 0}(1 + (z-1)V_i/H) \leq \exp(\frac{z-1}{H}\sum_{i \geq 0} V_i) \ . \tag{3.13}$$

Now we can re-write

$$\sum_{i \geq 0} V_i = \sum_{k=1}^{H} k\eta_k \ , \tag{3.14}$$

where η_k is the waiting time for the process V_i to go from k to $k-1$. The r.v.'s η_k are independent with (different) geometric distributions

$$\Pr\{\eta_k = u\} = (1 - k/H)^{u-1}\ k/H,\ u \geq 1.$$

(This is just the elementary argument for the classical coupon collector's problem with equally-likely coupons.) The associated generating function may be written as

$$E\theta^{\eta_k} = (1 - \frac{H}{k}(1-\theta^{-1}))^{-1} \ , \quad |\theta| < \frac{H}{H-k}. \tag{3.15}$$

Combining (3.13)–(3.15) gives

$$(ED_z)^{-1} \geq \prod_{k=1}^{H}(1 - \frac{H}{k}(1 - \exp(-\frac{(z-1)k}{H}))).$$

Now $1 - y^{-1}(1 - e^{-ay}) \geq 1 - a$ for $a, y > 0$, and so each term in the product is $\geq 1 - (z-1)$. This gives

$$ED_z \leq (2-z)^{-H},\ 1 < z < 2\ ,$$

and by Lemma 11 we obtain

$$\Pr\{B \geq b\} \leq 2\sqrt{\inf_{1<z<2} z^{-b}(2-z)^{-H}}\ .$$

Elementary calculus gives the exact infimum at $z = 2b/(b+H)$. Since the proof requires $z > 1$, the Lemma only holds for $b > H$. Theorem 3(i) indicates that lower values of b are not interesting anyway. ■

4. CONCLUDING REMARKS

Our main results of this paper concern *hashing with lazy deletion*. In particular, we assessed the average wasted space, the (maximum) excess space, the maximum space required by HwLD, etc. Our approach in Section 3 can be extended to evaluate data structures such as lists, dictionaries, stacks, priority queues and sweepline structures for geometry and VLSI [20].

There is an important conceptual difficulty buried beneath our asymptotics. Consider again a single $M|G|\infty$ with the queue length denoted by $N(t)$ and the arrival rate by ρ. Consider the behavior of the maximum $M_T(\rho) \equiv \max_{0\le t\le T} N(t)$. Having in mind the application to computer storage and VLSI, it is natural to suppose that ρ is at least moderately large. Theorem 6 says

$$M_T(\rho) \sim \frac{\log T}{\log\log T} \quad \text{as } T \to \infty,\ \rho \text{ fixed} .$$

It is natural to interpret this as establishing an approximation

$$M_T(\rho) \approx \frac{\log T}{\log\log T} \quad \text{for } T \text{ large} . \tag{4.1}$$

But substituting $T = e^{\rho}$ would give $M_T(\rho) \approx \rho/\log\rho$, which is absurd, because trivially $EM_T(\rho) \ge \rho$. Thus if $\rho = 100$ then e^{100} arrivals is not large enough for the asymptotics to be valid! (In fact, a little more analysis shows that the approximation (4.1) is valid asymptotically as $T, \rho \to \infty$ *if and only if* T increases super-exponentially fast in ρ.)

For practical applications, it is much more sensible to consider T being at most polynomially large in ρ. Classical queueing theory has apparently paid no attention to polynomial-time maxima. Mathieu and Vitter [17, 18] have initiated that type of analysis for hashing with lazy deletion. We have also obtained some results of this nature, which we may present in a forthcoming paper. Let us mention one result for the $M|G|\infty$ queue, stronger than those in [17, 18]. Qualitatively, the idea is that for large ρ, the standardized process $Y(t) \equiv (N(t) - \rho)/\rho^{1/2}$ behaves like the standardized Ornstein-Uhlenbeck process. So, appealing to the known asymptotic behavior of the Ornstein-Uhlenbeck process, it is easy to get a heuristic approximation in the spirit of [4, 14]

$$\Pr\{(M_T(\rho) - \rho)/\rho^{1/2} \le b\} \approx \exp(-Tb\phi(b)),\ b > 1 , \tag{4.2}$$

where $\phi(\cdot)$ is the *Standard Normal* density function. Proving rigorously the sharp result asserted in (4.2) seems difficult for technical reasons: the usual formalization via weak convergence of processes gives this result only for $T = T(\rho) \to \infty$ slowly with ρ. On the other hand, a crude consequence of (4.2) is

$$M_T(\rho) \sim \rho + \rho^{1/2}\psi(1/T)$$

where $\psi(\cdot)$ is the inverse function of $x\phi(x), x > 1$. The first term in the expansion of $\psi(1/T)$ is $\sqrt{2\log T}$. Thus (4.2) would imply the much weaker result: for T polynomially large in ρ,

$$M_T(\rho) \approx \rho + \rho^{1/2}(\sqrt{2\log T} + o(1)) \ .$$

This weaker result can be proved rigorously, under some additional assumptions.

APPENDIX

A. Proof of Lemma 7.

Lemma 7 asserted that $q(a) \to 1$ as $a \to \infty$, where $q(a)$ is the chance that a typical upcrossing time of X_t is a clearing time of X_t^*. Call an upcrossing of X_t at t *special* if no item present at time t was present at the previous upcrossing. Clearly a special upcrossing time is a clearing time for X_t^*, so it suffices to prove

$$q'(a) \equiv \text{ chance that a typical upcrossing is not special } \to 0 \text{ as } a \to \infty.$$

It is conceptually easier to reverse time and consider *downcrossings* from $a+1$ to a. A downcrossing at t is special iff all the items present at t have departed before the next downcrossing. Thus a sufficient condition for a downcrossing, at $t = 0$ without loss of generality, to be special is that the queue length does not return to $a+1$ before the time L at which every item present at time 0 has departed. So

$$q'(a) \leq \Pr_{\downarrow}\{X_t = a+1 \text{ for some } t < L\} \ ,$$

where $\Pr_{\downarrow}$ denotes probabilities for the stationary process conditional on there being a downcrossing from $a+1$ to a at time 0.

Write

$$X_t = X_t^+ + X_t^-, \quad t > 0$$

where X_t^+ denotes items in storage at time t which arrived after time 0, and X_t^- denotes items in storage at time t which arrived before time 0. We see that, under $\Pr_{\downarrow}$

(i) X_t^+ is the $M|G|\infty$ queue length process, started at 0;

(ii) $(a - X_t^-)$ is the Poisson counting process of rate $\theta(t) = \Pr\{S > t\}$, conditioned on the total number of events equaling a.

Now $q'(a)$ is further bounded by the sum of the following three probabilities.

$$\Pr_{\downarrow}\{X_t = a+1 \text{ for some } t \leq t_0\} \tag{A.1a}$$

$$\Pr\{X_t^+ > b(t) \text{ for some } t \geq t_0\} \tag{A.1b}$$

$$\Pr_{\downarrow}\{X_t^- \geq a + 1 - b(t) \text{ for some } t_0 \leq t < L\} \qquad (A.1c)$$

where t_0 and $b(t), t \geq 0$ are arbitrary.

Now it is easy to see that $a - X_t^- \to \infty$ as $a \to \infty$ with t fixed, and it follows that, for fixed t_0, the probability in (A.1a) tends to 0 as $a \to \infty$.

Next, if c_i is a non-decreasing integer-valued sequence with its continuous expansion equal to $c(t) \sim 2\log t/\log\log t$ then (using easy Poisson tail estimates) $\Pr\{X_i > c_i\} = o(i^{-2+\varepsilon})$, $\varepsilon > 0$. This leads to an estimate for the *stationary* process X

$$\Pr\{X_i \leq c_i \text{ for all sufficiently large } i\} = 1 \; .$$

Let A_i be the number of arrivals during the interval $[i, i+1]$. Since the (A_i) are independent Poissons, Corollary 5 shows

$$\Pr\{A_i \leq c_i' \text{ for all sufficiently large } i\} = 1$$

for suitable increasing integer-valued c_i'. Because $X_t \leq X_i + A_i$ on $i \leq t \leq i+1$, we deduce

$$\Pr\{X_t > b(t) \text{ for some } t \geq t_0\} \to 0 \text{ as } t_0 \to \infty$$

where $b(t) = c(\lfloor t \rfloor) + c'(\lfloor t \rfloor) \sim 3\log t/\log\log t$. (In fact, a more careful argument shows "3" could be replaced by "2".) Thus the quantity in (A.1b) tends to 0 as $t_0 \to \infty$, because X^+ is just the $M|G|\infty$ process started empty, and so we can take $X_t^+ \leq X_t$.

So we are left with the problem of bounding (A.1c): precisely, it suffices to prove

$$\lim_{t_0 \to \infty} \limsup_{a \to \infty} \Pr_{\downarrow}\{X_t^- \geq a + 1 - b(t) \text{ for some } t_0 \leq t < L\} = 0 \; . \qquad (A.2)$$

Since $X_t^- \geq 1$ for $t < L$, we may re-write the probability as

$$\Pr_{\downarrow}\{X_t^- \geq \max(1, a + 1 - b(t)) \text{ for some } t \geq t_0\}.$$

Consider the inverse function $b^{-1}(m) = \inf\{t : b(t) \geq m\}$. Since X_t^- is decreasing in t, it suffices to check the inequality for t of the form $b^{-1}(m)$, and the probability above

$$= \Pr_{\downarrow}\{X_{b^{-1}(m)}^- \geq a + 1 - m \text{ for some } b(t_0) \leq m \leq a\} \leq \sum_{m=b(t_0)}^{a} \frac{E_{\downarrow} X_{b^{-1}(m)}^-}{a+1-m} \; .$$

We now quote a standard fact. Let N_t be a Poisson counting process on $0 \leq t < \infty$ with rate $\theta(t)$ and such that $\int_0^\infty \theta(s)ds < \infty$. Then

$$E(N_t | N_\infty = a) = a \int_0^t \theta(s)ds \Big/ \int_0^\infty \theta(s)ds.$$

This follows from the fact ([19] exercise 2.24a) that, conditional on $N_\infty = a$, the positions of these a points are distributed as the positions of a points chosen independent from the distribution with distribution function $F(t) = \int_0^t \theta(s)ds / \int_0^\infty \theta(s)ds$. By (ii) above, we can apply this fact to $N_t = a - X_t^-$, to get

$$\begin{aligned} E_1 X_t^- &= a - a\int_0^t \theta(s)ds / \int_0^\infty \theta(s)ds \\ &= a\int_t^\infty \theta(s)ds / \int_0^\infty \theta(s)ds = aE(S-t)^+/ES. \end{aligned}$$

Thus the proof of (A.2) reduces to the proof of

$$\lim_{t_0\to\infty} \limsup_{a\to\infty} \sum_{m=b(t_0)}^{a} \frac{a}{a+1-m} E(S - b^{-1}(m))^+ = 0.$$

Splitting the sum at $a/2$, we see it is enough to prove

$$\sum_{m=1}^{\infty} E(S - b^{-1}(m))^+ < \infty$$

$$a \log a \quad E(S - b^{-1}(a/2))^+ \to 0 \text{ as } a \to \infty.$$

But these are simple consequences of the fact

$$\log b^{-1}(m) \sim \frac{1}{3} m \log m$$

together with the inequality

$$E(S-c)^+ \le \frac{ES\log^2 S}{\log^2 c},\ c > 1\ .$$

This completes the proof of Lemma 7. ∎

B. Proof of Lemma 9.

Lemma 9 asserted that for all sufficiently small $\theta > 0$,

$$E\exp(\theta \sum_{i=0}^{\infty} e^{-\beta\tau_i}) \le \exp(\theta\gamma).$$

Write $Q_n = \sum_{i=0}^{n} \exp(-\beta\tau_i)$. Then we have the recursion

$$Q_{n+1} \stackrel{d}{=} 1 + e^{-\beta\xi} Q_n$$

in which ξ and Q_n are taken independent. Consider some $\theta_0 > 0$. If we can show, by induction on n, that

$$E\exp(\theta Q_n) \le \exp(\theta\gamma) \quad \text{for all } \ 0 \le \theta \le \theta_0$$

then the Lemma 9 follows by letting $n \to \infty$. If the above holds for n then

$$\begin{aligned} E\exp(\theta Q_{n+1}) &= e^{\theta} E\exp(\theta e^{-\beta\xi} Q_n) \qquad \text{by our recurrence} \\ &\leq e^{\theta} E\exp(\theta e^{-\beta\xi}\gamma) \qquad \text{by induction assumption} \end{aligned}$$

To make the induction go through, we require this to be bounded by $\exp(\theta\gamma)$, and this rearranges to the requirement

$$E\exp(\theta(e^{-\beta\xi}\gamma - (\gamma - 1))) \leq 1, \qquad 0 \leq \theta \leq \theta_0. \tag{A.3}$$

But this fact (for some θ_0) follows from the choice of γ in relation (3.10b), which implies that $\frac{d}{d\theta}$(the left-hand side of equation (A.3)) at $\theta = 0$ is negative. ■

REFERENCES

[1] Abramowitz, M. and Stegun, I., *Handbook of Mathematical Functions,* Dover, New York (1964).

[2] Aven, O., Coffman, E.G. Jr. and Kogan, Y.A., *Stochastic Analysis of Computer Storage,* D. Reidel Publishing Company, (1987).

[3] Aho, A., Hopcroft, J. and Ullman, J., *Data Structures and Algorithms,* Addison-Wesley, Reading, MA (1983).

[4] Aldous, D., *Probability Approximations via the Poisson Clumping Heuristic,* Springer-Verlag, New York (1989)

[5] Anderson, C.W., Extreme value theory for a class of discrete distributions with applications to some stochastic processes, *J. Appl. Probability,* 7, 99-113 (1970).

[6] Berman, S., Limiting distribution of the maximum term in sequences of dependent random variables, *Ann. Math. Statist.,* 33 (1962).

[7] Durrett, R., *Probability: Theory and Examples,* Wadsworth, Belmont CA (1991).

[8] Feller, W., *An Introduction to Probability Theory and its Applications,* Vol. II, John Wiley & Sons, New York (1971).

[9] Flajolet, P., Francon, J. and Vuillemin, J., Sequence of operations analysis for dynamic data structures, *J. Algorithms,* 1, 111-141 (1980).

[10] Galambos, J., *The Asymptotic Theory of Extreme Order Statistics,* John Wiley & Sons, New York (1978).

[11] Kaplan, N., Limit Theorems for a $GI/G/\infty$ queue, *Ann. Probab.* 3, 780-789 (1975).

[12] Keilson, J., A limit theorem for passage times in ergodic regenerative processes, *Ann. Math. Statist.* 37, 866-870 (1966).

[13] Kleinrock, L., *Queueing Systems, Vol. 1,* John Wiley & Sons (1976).

[14] Louchard G., Kenyon, C. and Schott, R., Data Structures Maxima, Universite Libre de Bruxelles, TR.216 (1990).

[15] Louchard, G., Randrianarimanana, B. and Schott, R., Dynamic algorithms in D.E. Knuth's model: A probabilistic analysis, *Theoretical Computer Science*, to appear.

[16] Morrison, J., Shepp, L. and Van Wyk, C., A queueing analysis of hashing with lazy deletion, *SIAM J. Comput.,* 16, 1155-1164 (1987).

[17] Mathieu, C. and Vitter, J.S., Maximum queue size and Hashing with lazy deletion, Brown University, Technical Report (1987).

[18] Mathieu, C. and Vitter, J.S., General methods for the analysis of the maximum size of dynamic data structures, *Proc. ICALP'89*, 473-487, Stresa 1989.

[19] Ross, S.M., *Stochastic Processes*, John Wiley & Sons, Chichester (1983).

[20] Szymanski, T.G. and Van Wyk, C., Space efficient algorithms for VLSI artwork analysis, *Proc. 20th Design Automation Conference,* 734-739 (1983).

[21] Szpankowski, W., On the maximum queue length with applications to data structure analysis, *Proc. Twenty-Seventh Allerton Conference*, 263-272 (1989).

[22] Van Wyk, C. and Vitter, J.S., The complexity of hashing with lazy deletion, *Algorithmica*, 1, 17-29 (1986).

[23] Wolff, W., *Stochastic Modeling and the Theory of Queues*, Prentice-Hall, Englewood Cliffs (1989).