Geometric Generalizations of the Power of Two Choices

John W. Byers* Jeffrey Considine* Michael Mitzenmacher†

Abstract

A well-known paradigm for load balancing in distributed systems is the "power of two choices," whereby an item is stored at the less loaded of two (or more) random alternative servers. We investigate the power of two choices in natural settings for distributed computing where items and servers reside in a geometric space and each item is associated with the server that is its nearest neighbor. This is in fact the backdrop for distributed hash tables such as Chord, where the geometric space is determined by clockwise distance on a one-dimensional ring. Theoretically, we consider the following load balancing problem. Suppose that servers are initially hashed uniformly at random to points in the space. Sequentially, each item then considers d candidate insertion points also chosen uniformly at random from the space, and selects the insertion point whose associated server has the least load. For the one-dimensional ring, and for Euclidean distance on the two-dimensional torus, we demonstrate that when n data items are hashed to n servers, the maximum load at any server is $\log \log n / \log d + O(1)$ with high probability. While our results match the well-known bounds in the standard setting in which each server is selected equiprobably, our applications do not have this feature, since the sizes of the nearest-neighbor regions around servers are non-uniform. Therefore, the novelty in our methods lies in developing appropriate tail bounds on the distribution of nearest-neighbor region sizes and in adapting previous arguments to this more general setting. In addition, we provide simulation results demonstrating the load balance that results as the system size scales into the millions.

1 Introduction

A well-known paradigm for balancing load is the "power of two choices" [1, 9, 10, 15], whereby an item is stored at the less loaded of two (or more) random alternatives, which we refer to variously as bins and servers. These methods are used in standard hashing with chaining to reduce the maximum number of items, or load, in a bin with high probability. Using these methods, two or more hash functions are used to pick candidate bins for each item to be inserted. Prior to insertion, the loads of the candidate bins are compared and the item is inserted into the bin with the least load. Similarly, to search for an item, the hash functions are applied and each candidate bin is examined to locate the item.

We consider applying this load balancing paradigm in natural settings for distributed computing where the servers and items are placed in a geometric space, and an item is associated with the server that is its nearest neighbor. After an initial random placement of the servers, each item considers two (or more) random choices for its location; it picks the one where the nearest neighboring server has the least load. In contrast to a typical application of this paradigm, where each of the *servers* is selected uniformly at random, our application picks *locations* in the geometric space uniformly

*Boston University, Dept. of Computer Science. Supported in part by NSF grants ANI-0093296 and ANI-9986397. E-mail: {byers,jconsidi}@cs.bu.edu.

†Harvard University, Div. of Engineering and Applied Sciences. Supported in part by NSF grants CCR-9983832, CCR-0118701, CCR-0121154, and an Alfred P. Sloan Research Fellowship. E-mail: michaelm@eecs.harvard.edu.

at random. As a result, the probability of an item considering a given server s is non-uniform; rather, it is proportional to the volume of the geometric space that lies closest to s. For example, using Euclidean distance and considering the Voronoi diagram induced by the random placement of a set of servers on the plane, the probability that a randomly selected location is closest to s is proportional to the size of the Voronoi region of s.

Our contributions are threefold. First, we describe motivating applications which exemplify nearest-neighbor load balancing on the 1-dimensional ring and the 2-dimensional torus. Our second contribution, and our primary focus, is on theoretical foundations. We provide an analytical framework for reasoning about nearest-neighbor load balancing using the power of two choices paradigm and extend the seminal results of Azar, Broder, Karlin and Upfal [1]. Their result holds in the standard setting where all bins are equally likely to be selected by each hash function; they show in this case that if there are n items, n bins, and $d \geq 2$ hash functions, the maximum load of any bin is only $\log\log n/\log d + O(1)$ with high probability [1]. More generally, when using two choices, the maximum load is more tightly concentrated around the mean than when no choices are available. Using techniques which are interesting in their own right, we demonstrate that these results still hold for nearest-neighbor load balancing on the ring and on the 2-dimensional torus. Our results in one dimension necessitate bounding the number of arcs of a given length appropriately; similarly, our results in two dimensions rely on bounding the number of Voronoi regions of a given area when the points inducing the Voronoi diagram are placed uniformly at random. Finally, our third contribution is to present experimental results demonstrating the effectiveness of the resulting load balance in practice as the size of the system scales up to millions of items and servers.

1.1 Applications

Our original motivation arises from recent work on distributed hash tables (DHTs), which have been proposed as a fundamental building block for peer-to-peer systems [8, 13, 12, 14, 16]. In a standard consistent hashing approach for constructing DHTs, both servers and keys are hashed onto a one dimensional ring. Keys are then assigned to the nearest server in the clockwise direction. Searching is enabled through the addition of server-to-server connections. To ensure connectivity, the ring structure is embedded in the overlay, by having servers connect to their direct neighbors in the ring. Additional overlay edges spanning larger arcs around the ring facilitate fast searches. For example, in Chord [14], each server maintains a carefully constructed "finger table" of connections to other servers. These tables are each of logarithmic size and enable searches which are routed through only a logarithmic number of servers.

The naive implementation of consistent hashing described above can result in significant load imbalance, due to the non-uniformity of arc lengths associated with each server. The authors of Chord propose a more sophisticated solution whereby each server simulates a logarithmic number of "virtual servers," thus assigning each server several smaller segments whose total size is more tightly concentrated around the expectation. However, as we demonstrate using simulations in [3], a simpler and more cost-effective solution is to apply the power of two choices paradigm described above.[1] In this context, the geometric space is the ring, and after servers are initially placed uniformly at random, items simply consider two (or more) prospective locations on the ring at random and choose the one which maps to the least loaded server. Simple refinements to the Chord lookup procedure based on this approach are detailed in [3].

[1]To aid the PC, we would like to clarify for the committee reviewers that an initial short summary detailing the application of two choices to consistent hashing on the ring has been accepted to the IPTPS 2003 workshop. A version of this summary can be found at [3]. This five-page workshop summary has not been published in a conference proceedings; moreover, the current submission is substantially different from this summary (which focuses exclusively on one peer-to-peer application and contains none of the proofs, nor covers the case of the 2-D torus). We strongly believe that our PODC submission merits its own conference publication; we hope the committee will agree.

For a 2-dimensional example of a situation where nearest-neighbor load balancing might apply, suppose that one's bank wanted to try to balance the load among its automatic teller machines throughout the city. For each customer, it suggests a base machine, which will be the closest machine to either the customer's home or work location. Assuming machines and customer locations are distributed randomly throughout the city, our 2-dimensional balls and bins framework models the sequential assignment of customers to teller machines.[2] We expect the fact that the power of two choices applies in these more general geometric settings will prove useful for other distributed applications.

2 The Layered Induction Argument for Random Circular Arcs

In this section, we prove our first main result, providing the key lemmas needed in order to extend the original argument of Azar et al. to the setting where bins correspond to arcs generated by placing n points randomly on the circle.

Theorem 1 *Suppose that n points are placed independently and uniformly at random on the boundary of a circle with circumference 1. The n induced arcs correspond to bins. Now balls are placed sequentially into the n bins as follows. Each ball chooses d points independently and uniformly at random from the boundary of the circle; each of the d points corresponds to the bin whose arc contains that point. The ball is placed into the least full bin of these d bins at the time of the placement, with ties broken arbitrarily. After all the balls are placed, with probability $1 - o(1/n)$ the number of balls in the fullest bin is at most $\log \log n / \log d + O(1)$.*

The proof is fairly technical, although it generally follows the proof of Azar et al. Here, we sketch the proof, highlighting the points where our analysis differs and motivating the key lemmas we prove in this section. For any given i, instead of trying to determine the number of bins with load *exactly* i, it is easier to study the number of bins with load *at least* i. Let the *height* of a ball be one more than the number of balls already in the bin in which the ball is placed. That is, if we think of balls as being stacked in the bin by order of arrival, the height of a ball is its position in the stack. Suppose we know that the number of bins with load at least i, over the entire course of the process, is bounded above by β_i. We wish to find a β_{i+1} such that, with high probability, the number of bins with load at least $i+1$ is bounded above by β_{i+1} over the course of the entire process with high probability. We find an appropriate β_{i+1} by bounding the number of balls of height at least $i+1$, which gives an upper bound for the number of bins with at least $i+1$ balls.

A ball has height at least $i+1$ only if each of its d bin choices have load at least i. If all the bins were equally likely to be chosen, then conditioned on the number of bins of height i being at most β_i, the probability that each choice yields a bin of load at least i is at most $\frac{\beta_i}{n}$. Therefore the probability a ball would have height $i+1$ would be $\left(\frac{\beta_i}{n}\right)^d$, and a Chernoff bound would assure that the number of balls of height at least $i+1$ would be at most $2\left(\frac{\beta_i}{n}\right)^d$ with high probability. This is the fact used in the original proof, yielding the the recursion $\beta_{i+1} \leq 2\left(\frac{\beta_i}{n}\right)^d$, which gives the upper bound.

In our setting, the at most β_i bins with load at least i might correspond to bins with arc lengths larger than average. Hence, in order to bound the probability that a ball has height at least $i+1$, we bound for the total arc length of the β_i longest bins. Using this bound, we bound the probability

[2]In practice, the distribution of ATMs and customers may be highly non-uniform. While we only prove bounds on the load for the uniform case, experience suggests that the power of two choice methodology will nevertheless often work well at reducing the maximum load in other situations.

a ball has height $i+1$, and derive an appropriate recursion for β_{i+1}. This recursion, which is somewhat more complicated than in the original argument, gives the $\log\log n/\log d + O(1)$ bound.

Clearly a key step is to bound the total length of the β_i longest arcs. We accomplish this using Chernoff bounds. Specifically, let $B(n,p)$ be a Bernoulli random variable with parameters n and p. We make repeated use of the following form of Chernoff's bound.

Lemma 2 *(Chernoff's bound)*

$$\mathbf{Pr}(B(n,p) \geq 2np) \leq \mathrm{e}^{-np/3}.$$

We use the Chernoff bound to bound the number of arcs of length at least x for various values of x. Unfortunately, the arc lengths are *dependent* random variables, so the Chernoff bound cannot immediately be applied. One way to cope with this dependence is to use a martingale argument. While the obvious such argument (which for completeness we give below) gives a weaker bound, it would suffice for our main result. However, we have found that in the case of arc lengths that we can make use of *negative dependence* [5]. Since this appears interesting in its own right, and slighly simplies Theorem 1, we give the details.

For the purposes of this paper, we say that a collection of 0-1 random variables $X_1, X_2, \ldots, X_n$ is *negatively dependent* if and only if

$$\mathbf{E}\left[X_{i_1}X_{i_2}\cdots X_{i_k}\right] \leq \mathbf{E}\left[X_{i_1}\right]\mathbf{E}\left[X_{i_2}\right]\cdots\mathbf{E}\left[X_{i_k}\right]$$

for any distinct indices $i_1, i_2, \ldots, i_k$ in the range $[1,n]$. Let $X = \sum_{i=1}^{n} X_i$. A simple consequence of the above fact is then, since each variable takes only the values 0 or 1,

$$\mathbf{E}\left[\mathrm{e}^{tX}\right] \leq \prod_{i=1}^{n}\mathbf{E}\left[\mathrm{e}^{tX_i}\right].$$

(This can be seen by expanding e^{tX} using the Taylor series expansion, and using $\mathbf{E}\left[X_{i_1}^{j}\right] = \mathbf{E}\left[X_{i_1}\right]$ since all variables are 0-1.) This is all that is required for the Chernoff bound for the upper tail to hold, following the standard proof [11]. Hence, we can apply the Chernoff bound of Lemma 2 as long as our random variables are negatively dependent.

We now prove some key lemmata regarding the distribution of the arc lengths when n points are thrown randomly onto the circle, as well as negative dependence between sufficiently long arc lengths.

Lemma 3 *Suppose n points are thrown independently and uniformly at random on the boundary of the unit circle. Let $Z_j = 1$ if the counterclockwise arc from the jth point has length at least c/n, and $Z_j = 0$ otherwise. Then the Z_j are negatively dependent.*

Proof: Without loss of generality, consider the k random variables $Z_1, Z_2, \ldots, Z_k$. We have $\mathbf{E}[Z_i] = (1-\frac{c}{n})^{n-1}$ for all i, and hence

$$\prod_{i=1}^{k}\mathbf{E}\left[Z_i\right] = \left(1-\frac{c}{n}\right)^{k(n-1)}.$$

Now $Z_1Z_2\ldots Z_k = 1$ if and only if all of the arcs associated with each of the k points has length at least c/n; we bound the probability of this happening. Consider the points being placed one at a time. We first consider the first k points, which require some care.

For $1 \leq i \leq k$, Let A_i be the event that the first i points each have non-overlapping arcs of length at least c/n. Let B_i be the event that the ith point does not fall within c/n (in counterclockwise distance) of any of the first $(i-1)$ points. Let C_i be the event that the arc of length c/n from the ith point does not contain any of the first $(i-1)$ points. Clearly, $\mathbf{Pr}(A_1) = 1$, and

$$\mathbf{Pr}(A_{i+1}) \quad = \quad \mathbf{Pr}(A_i) \cdot \mathbf{Pr}(B_{i+1} \mid A_i) \cdot \mathbf{Pr}(C_{i+1} \mid A_i B_{i+1}).$$

The probability $\mathbf{Pr}(B_{i+1} \mid A_i)$ is easy to calculate as the arcs are necessarily disjoint.

$$\mathbf{Pr}(B_{i+1} \mid A_i) = 1 - \frac{ci}{n} \leq \left(1 - \frac{c}{n}\right)^i.$$

To compute $\mathbf{Pr}(C_{i+1} \mid A_i B_{i+1})$, we examine an equivalent view of the experiment. Consider starting with a circle of circumference $1 - i(c/n)$ and placing the $i+1$ points uniformly at random. Then think of expanding each of the first i points into an empty arc of length c/n in the counterclockwise direction, giving a corresponding circle of length 1 satisfying events A_i and B_{i+1}. The distribution of point placements we obtain in this fashion is equivalent to that of the original placement approach when conditioned on events A_i and B_{i+1}. The advantage of this point of view is that if we think of the $(i+1)$st point as being placed first, then it is clear that

$$\mathbf{Pr}(C_{i+1} \mid A_i B_{i+1}) = \left(\frac{1 - c(i+1)/n}{1 - ci/n}\right)^i \leq \left(1 - \frac{c}{n}\right)^i.$$

Combining the above, we find that the probability that after the placement of the first k points there are k disjoint arcs of length at least c/n is at most

$$\prod_{i=1}^{k-1} \mathbf{Pr}(B_{i+1} \mid A_i) \cdot \mathbf{Pr}(C_{i+1} \mid A_i B_{i+1}) \leq \prod_{i=1}^{k-1} \left(1 - \frac{c}{n}\right)^{2i} = \left(1 - \frac{c}{n}\right)^{k^2 - k}.$$

For the remaining $n-k$ points, the probability that they miss the k disjoint arcs of length c/n corresponding to the first k points is

$$\left(1 - \frac{ck}{n}\right)^{n-k} \leq \left(1 - \frac{c}{n}\right)^{k(n-k)}.$$

Hence the probability that $Z_1 Z_2 \ldots Z_k = 1$, or $\mathbf{E}[Z_1 Z_2 \ldots Z_k]$, is bounded above by

$$\left(1 - \frac{c}{n}\right)^{k(n-k)} \left(1 - \frac{c}{n}\right)^{(k^2-k)} = \left(1 - \frac{c}{n}\right)^{k(n-1)},$$

proving the negative dependence. ■

Lemma 4 *Let N_c be the number of arcs of length at least c/n. For $n \geq c \geq 2$,*

$$\mathbf{Pr}(N_c \geq 2n\mathrm{e}^{-c}) \leq \mathrm{e}^{-n\mathrm{e}^{-c}/3}.$$

Proof:

Again let $Z_j = 1$ if the jth point is the starting endpoint of an arc of length at least c/n, and 0 otherwise. From the result in Lemma 3, there is negative dependence between the Z_j's; therefore, Chernoff bounds apply. The probability $Z_j = 1$ can be bounded above by

$$\left(1 - \frac{c}{n}\right)^{n-1} \leq \mathrm{e}^{-c}$$

where the inequality holds for $c \geq 2$. Hence $\mathbf{E}[N_c] = \sum_{j=1}^{n} Z_j \leq n\mathrm{e}^{-c}$.

Applying the standard Chernoff bound of Lemma 2 yields

$$\mathbf{Pr}(N_c \geq 2n\mathrm{e}^{-c}) \leq \mathrm{e}^{-n\mathrm{e}^{-c}/3}.$$

■

For other variations of the problem (including our application to the the torus in Section 3), we have not yet been able to show that negative dependence applies. For completeness, we demonstrate a weaker bound on the number of arcs exceeding a given length via a simple martingale argument. This bound would be sufficient for Theorem 1, although the proof would have to modified slightly to make up for the weaker bound.

Lemma 5 *Let N_c be the number of arcs of length at least c/n. For $n \geq c \geq 2$,*

$$\mathbf{Pr}(N_c \geq 2n\mathrm{e}^{-c}) \leq \mathrm{e}^{-n\mathrm{e}^{-2c}/8}.$$

Proof: Consider the n random points being placed on the circle one at a time. Let X_i be the location of the ith point, and $Y_i = \mathbf{E}[N_c \mid X_1, X_2, \ldots, X_i]$. Then the sequence of Y_i is a Doob martingale with expectation $\mathbf{E}[N_c]$. We have seen in Lemma 4 that $\mathbf{E}[N_c] \leq n\mathrm{e}^{-c}$.

The Doob martingale satisfies the Lipschitz condition for the independent random variables $X_1, X_2, \ldots, X_n$, with a Lipschitz bound of 2. That is, changing the value of one of the X_i (that is, the placement of any of the n points) arbitrarily can change the total number of arcs of length at least c/n by at most 2. This is because each additional point can only increase or decrease N_c by at most 1, either by splitting an arc of length at least c/n into two smaller arcs both of length at least c/n or by splitting an arc of length at least c/n into two smaller arcs both smaller than c/n. Similarly, removing a single point changes N_c by at most 1.

Hence, applying (a one-sided) Azuma's inequality (see, e.g., [11]), we have

$$\mathbf{Pr}(N_c \geq 2n\mathrm{e}^{-c}) \leq \mathrm{e}^{-n\mathrm{e}^{-2c}/8}.$$

■

Using Lemma 4, we obtain a very loose bound on the total length of the a longest arcs when n points are randomly thrown on the circle. We note that the bound below is not optimized; it is simply sufficient for our purposes.

Lemma 6 *Let $(\ln n)^2 \leq a \leq n/64$. Then the probability that the sum of the length of the longest a arcs exceeds $2\frac{a}{n}\ln\frac{n}{a}$ is $o(1/n^2)$.*

Proof: Consider the largest a arcs. We first bound the length of the smallest $a/2$ of these a arcs. We apply the argument recursively to obtain the final bound, using some care for the final tail.

From Lemma 4, the probability there are more than $2n\mathrm{e}^{-c}$ arcs of length at least c/n is at most $\mathrm{e}^{-n\mathrm{e}^{-c}/3}$ when $c \geq 2$. Rephrasing this statement by letting $b/2 = 2n\mathrm{e}^{-c}$, the probability that there are more than $b/2$ arcs of length $\ln\frac{4n}{b}/n$ is at most $\mathrm{e}^{-b/12}$ for $(\ln n)^2/64 \leq b \leq n/64$. Hence, the probability that the sum of the length of the smallest $b/2$ of the longest b arcs exceeds $\frac{b}{2n}\ln\frac{4n}{b}$ for b in the above range is at most $\mathrm{e}^{-b/12}$.

We also use the additional fact that the probability that the longest arc exceeds $4\ln n/n$ is

$$n\left(1 - \frac{4\ln n}{n}\right)^{n-1} \leq \frac{1}{n^3}.$$

Let j be the smallest integer such that $a/2^j \leq (\ln n)^2/32$ for a in the range given by the theorem. Using the above facts, we have that the sum of the lengths of the largest a arcs exceeds

$$\sum_{k=0}^{j} \frac{a}{2^{k+1}n}\ln\left(\frac{4n}{a/2^k}\right) + \frac{(\ln n)^2}{64}(4\ln n)$$

with probability at most

$$\sum_{k=0}^{j} e^{-(a/2^k)/12} + \frac{1}{n^3}.$$

This probability is $o(1/n^2)$, and the sum is bounded above by $2\frac{a}{n}\ln\frac{n}{a}$ for a in the given range and n suitably large as follows:

$$\begin{aligned}
\sum_{k=0}^{j} \frac{a}{2^{k+1}n}\ln\left(\frac{4n}{a/2^k}\right) + \frac{(\ln n)^2}{64}\frac{4\ln n}{n} &\le \sum_{k=0}^{\infty}\frac{a}{2^{k+1}n}\ln\frac{4n}{a} + \sum_{k=0}^{\infty}\frac{a}{2^{k+1}n}\ln 2^k + \frac{(\ln n)^3}{16n} \\
&\le \frac{a}{n}\ln\frac{4n}{a} + \frac{a\ln 2}{2n}\sum_{k=0}^{\infty}\frac{k}{2^k} + \frac{(\ln n)^3}{16n} \\
&\le \frac{a}{n}\ln\frac{n}{a} + \frac{a}{n}3\ln 2 + \frac{(\ln n)^3}{16n} \\
&\le 2\frac{a}{n}\ln\frac{n}{a}.
\end{aligned}$$

■

We now prove Theorem 1. Again, the structure of the proof follows that of Azar et al. with the addition of the requirement that non-uniform arc lengths must be handled. We use the following notation: the state at time t refers to the state of the system immediately after the tth ball is placed. The variable $h(t)$ denotes the height of the tth ball, and $\nu_i(t)$ and $\mu_i(t)$ refer to the number of bins with load at least i and the number of balls with height at least i at time t, respectively. We use ν_i and μ_i for $\nu_i(n)$ and $\mu_i(n)$ when the meaning is clear.

We make use of the following lemma.

Lemma 7 *Let $X_1, X_2, \ldots, X_n$ be a sequence of random variables in an arbitrary domain, and let $Y_1, Y_2, \ldots, Y_n$ be a sequence of binary random variables, with the property that $Y_i = Y_i(X_1, \ldots, X_i)$. If*

$$\mathbf{Pr}(Y_i = 1 \mid X_1, \ldots, X_{i-1}) \le p,$$

then

$$\mathbf{Pr}\left(\sum_{i=1}^{n} Y_i \ge k\right) \le \mathbf{Pr}(B(n,p) \ge k).$$

Proof: Following the earlier sketch, we shall construct values β_i so that $\nu_i(n) \le \beta_i$, for all i, with high probability. We emphasize that no effort has been made to optimize the $O(1)$ constant given in the proof.

Let $\beta_{256} = \frac{n}{256}$, and

$$\beta_{i+1} = 2n\left(2\frac{\beta_i}{n}\ln\frac{n}{\beta_i}\right)^d, \tag{1}$$

for $256 \le i < i^*$, where i^* is to be determined, but will always be $O(\log\log n)$. We let $\mathcal{E}_i$ be the event that $\nu_i(n) \le \beta_i$. Note that $\mathcal{E}_{256}$ holds with certainty. We now show that, with high probability, if $\mathcal{E}_i$ holds then $\mathcal{E}_{i+1}$ holds, for $256 \le i \le i^* - 1$.

To begin, we start by noting that we implicitly condition in all probabilistic statements that follow subsequently (until the end of the proof) that Lemma 6 holds whenever $a = \beta_i$ for $256 \le i \le i^* - 1$. That is, we assume that initially the points to create the arcs are selected, and that after that initial process the sum of the longest β_i arc lengths are bounded correctly as in Lemma 6, and further that the longest arc length is at most $4\ln n/n$ (as used in the Lemma). By the union bound, and the fact that i^* is $O(\log\log n)$, the probability that this assumption does not hold is bounded above is $o(1/n)$, and as we explain at the end of the argument, this does not therefore affect our final result that the probability the upper bound fails to hold is only $o(1/n)$.

Now fix a value of i in the given range. Let Y_t be a binary random variable such that

$$Y_t = 1 \textbf{ iff } h(t) \geq i+1 \textbf{ and } \nu_i(t-1) \leq \beta_i.$$

That is, Y_t is 1 if the height of the tth ball is at least $i+1$ and at time $t-1$ there are fewer than β_i bins with load at least i. Notice that for the tth ball to have height at least $i+1$, all d of its choices for bins must have at least i balls. If there are at most β_i bins with at least i balls, the total arc length of the bins with at least i balls is at most $2\frac{\beta_i}{n}\ln\frac{n}{\beta_i}$ by our previous assumption, as long as $(\ln n)^2 \leq \beta_i \leq n/256$.

We can conclude the following. Let ω_j represent the bins selected by the jth ball. Then

$$\mathbf{Pr}(Y_t = 1 \mid \omega_1, \ldots, \omega_{t-1}) \leq \left(2\frac{\beta_i}{n}\ln\frac{n}{\beta_i}\right)^d \stackrel{\text{def}}{=} p_i.$$

Thus, from Lemma 7, we may conclude that

$$\mathbf{Pr}(\textstyle\sum_{i=1}^n Y_t \geq k) \leq \mathbf{Pr}(B(n,p_i) \geq k).$$

Conditioned on $\mathcal{E}_i$, we have $\sum Y_t = \mu_{i+1}$. Thus

$$\begin{aligned}
\mathbf{Pr}(\nu_{i+1} \geq k \mid \mathcal{E}_i) &\leq \mathbf{Pr}(\mu_{i+1} \geq k \mid \mathcal{E}_i) \\
&= \mathbf{Pr}(\textstyle\sum Y_t \geq k \mid \mathcal{E}_i) \\
&\leq \frac{\mathbf{Pr}(\sum Y_t \geq k)}{\mathbf{Pr}(\mathcal{E}_i)} \\
&\leq \frac{\mathbf{Pr}(B(n,p_i) \geq k)}{\mathbf{Pr}(\mathcal{E}_i)}
\end{aligned}$$

Letting $k = \beta_{i+1}$ in the above, we have that

$$\mathbf{Pr}(\nu_{i+1} \geq \beta_{i+1} \mid \mathcal{E}_i) \leq \frac{\mathbf{Pr}(B(n,p_i) \geq 2np_i)}{\mathbf{Pr}(\mathcal{E}_i)} \leq \frac{\mathrm{e}^{-np_i/3}}{\mathbf{Pr}(\mathcal{E}_i)}.$$

We conclude that

$$\mathbf{Pr}(\neg\mathcal{E}_{i+1} \mid \mathcal{E}_i) \leq \frac{1}{n^2\mathbf{Pr}(\mathcal{E}_i)}$$

whenever $p_i \geq 6\ln n/n$. Using the bound

$$\mathbf{Pr}(\neg\mathcal{E}_{i+1}) \leq \mathbf{Pr}(\neg\mathcal{E}_{i+1} \mid \mathcal{E}_i)\mathbf{Pr}(\mathcal{E}_i) + \mathbf{Pr}(\neg\mathcal{E}_i),$$

we have

$$\mathbf{Pr}(\neg\mathcal{E}_{i+1}) \leq \frac{1}{n^2} + \mathbf{Pr}(\neg\mathcal{E}_i).$$

To finish, let i^* be the smallest value of i for which $p_i < 6\ln n/n$. We need that i^* is in fact $\frac{\log\log n}{\log d} + O(1)$. Proving this requires substantially more technical effort than the similar step for the original proof of Azar et al. [1], as the recursion (1) is more complex. The details of the technical manipulation appear in the appendix.

Following the previous line of reasoning,

$$\mathbf{Pr}(\nu_{i^*+1} \geq 12\ln n \mid \mathcal{E}_{i^*}) \leq \frac{\mathbf{Pr}(B(n, 6\ln n/n) \geq 12\ln n)}{\mathbf{Pr}(\mathcal{E}_{i^*})} \leq \frac{1}{n^2\mathbf{Pr}(\mathcal{E}_{i^*})},$$

and so

$$\mathbf{Pr}(\nu_{i^*+1} \geq 12 \ln n) \leq \frac{1}{n^2} + \mathbf{Pr}(\neg \mathcal{E}_{i^*}).$$

Further, under the explicit assumption that the maximum length arc is at most length $4 \ln n/n$,

$$\begin{aligned}\mathbf{Pr}(\mu_{i^*+2} \geq 2 \mid \nu_{i^*+1} < 12 \ln n) &\leq \frac{\mathbf{Pr}(B(n, (48(\ln n)^2/n)^d) \geq 2)}{\mathbf{Pr}(\nu_{i^*+1} < 12 \ln n)} \\ &\leq \frac{\binom{n}{2}(48(\ln n)^2/n)^{2d}}{\mathbf{Pr}(\nu_{i^*+1} < 12 \ln n)}\end{aligned}$$

by a simple union bound, and thus

$$\mathbf{Pr}(\mu_{i^*+2} \geq 2) \leq \binom{n}{2}(48(\ln n)^2/n)^{2d} + \mathbf{Pr}(\nu_{i^*+1} \geq 12 \ln n).$$

Combining our bounds, we find that conditioned on the arc lengths being appropriately distributed as per Lemma 6,

$$\mathbf{Pr}(\mu_{i^*+2} \geq 2) \leq \binom{n}{2}(48(\ln n)^2/n)^{2d} + \frac{i^*+1}{n^2} = o(1/n).$$

As this condition holds with probability $1 - o(1/n)$, $\mathbf{Pr}(\mu_{i^*+2} \geq 2) = o(1/n)$ even without this condition. Hence $\mathbf{Pr}(\mu_{i^*+2} \geq 2) = 1 - o(1/n)$. But if $\mu_{i^*+2} \leq 1$, the maximum load is $i^* + 2$, which is $\log \log n / \log d + O(1)$, proving the theorem. ■

We make several remarks regarding this result. First, the $O(1)$ constant chosen is excessive for practical considerations, and could easily be improved with some further technical work. Second, there is a corresponding $\log \log n / \log d - O(1)$ lower bound, which follows immediately from the lower bound for the standard case studied by Azar, Broder, Karlin, and Upfal [1]. Third, this proof can be extended to the case where the number of balls is $m \neq n$. Following the argument [1], we find the maximum load is $O(m/n) + O(\log \log n / \log d)$ with high probability. Stronger bounds, following the lines of [2], may be possible. Fourth, Vöcking suggested a variation of the original scheme that achieves an improved upper bound of $\log \log n / d \log \phi_d + O(1)$ by breaking ties in a novel manner [15]. In this setting, the variation would correspond to each ball picking one point uniformly from each of the d intervals $[0, i/d)$ for $i = 1, 2, \ldots, d$, going to the corresponding least loaded bin, and breaking ties toward the interval with the lowest corresponding value of i. Our proof could be modified to show that the $\log \log n / d \log \phi_d + O(1)$ upper bound holds in this setting with this variation as well. In our experimental section, we describe an alternative tie breaking scheme that appears to provide slightly better performance than even this scheme.

3 Two Choices in Voronoi Diagrams

In this section, we demonstrate another specific case where two choices gives a similar benefit in load balancing on a distributed system with a non-trivial underlying geometry. Instead of servers being points on the boundary of a circle, let servers correspond to points placed uniformly at random on the two-dimensional unit torus.[3] That is, we work in the space of points (x, y) with $0 \leq x, y \leq 1$, with wraparound along both axes. Sequentially, items are hashed to d candidate insertion locations

[3] Our results can be made to apply to the unit circle or unit square; we use the torus as the symmetry avoids the technicalities introduced by boundaries. Also, as should become clear, our argument generalizes to higher constant dimension.

on the torus. Each of these locations is associated with the corresponding nearest server, and the item chooses the insertion point where the corresponding server has the least load. Thus, in terms of the Voronoi diagram associated with the servers, each server is ultimately responsible for all items inserted into its region in the Voronoi diagram. One can show that the largest Voronoi region when there are n servers has area $\Theta(\log n)$ with high probability, similar to the case of arcs on the circle. Hence when $d = 1$ it is again easy to show that some server is responsible for $\Theta(\log n)$ items with high probability. With $d \geq 2$ choices, the maximum load when n items are hashed is again $\log\log n/\log d + O(1)$.

The key here is to again show that the distribution of the area of Voronoi regions is close enough to the uniform distribution that the argument of Theorem 1 applies. That is, we need a lemma of the same form as Lemma 5. There has been a fair amount of work on the distribution of the area of regions in random Voronoi diagrams, particularly in two dimensions; see, for example, the work by Miles [6] or Moller [7]. Knowing the distribution is insufficient for our purposes; however, since the areas of the Voronoi regions are potentially dependent in non-trivial ways. Interestingly, we have not found tail bounds of the sort we need in previous literature.

It seems intuitive that Voronoi region sizes would be negatively dependent. We have not found a proof of this in the literature, nor have we been able to prove it ourselves. But by making use of an appropriate set of random variables, we can develop simple bounds via martingales that prove effective for this setting. We begin with a key geometric lemma. Consider any point u in the torus. Let the circular area of area c/n around u be divided into six subregions, as in Figure 1, each of size $c/6n$. Specifically, taking 0 degrees to be parallel to the x-axis and to the right of u, the first subregion is the area from 0 to 60 degrees, and so on.

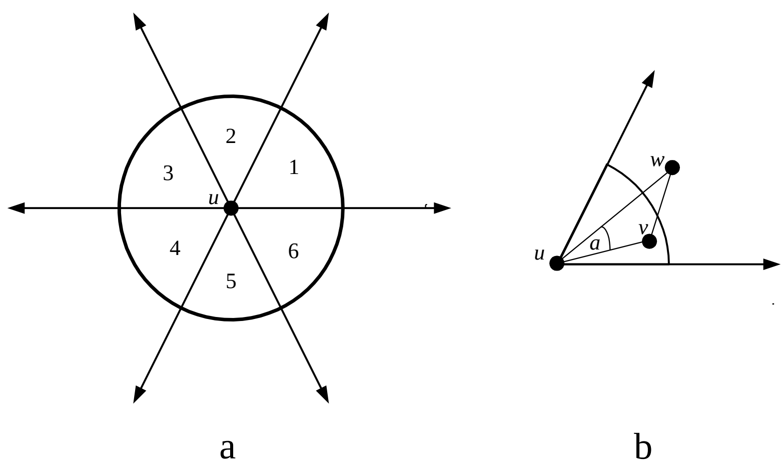

Figure 1: (a) 6 subregions of a circular region around a point u, each corresponding to a 60 degree arc. (b) If v lies in subregion 1 of u, then v is closer to w than u.

Lemma 8 *In a Voronoi diagram derived from n points, if the area of a Voronoi region for a point u is at least c/n, then at least one of the six subregions around u described above does not contain another of the $n-1$ points.*

Proof: Suppose without loss of generality the first subregion of area $c/6n$ around a point u has another of the other $n-1$ randomly placed points, call it v, lying within it. Then we claim that any point in the torus making an angle of between 0 and 60 degrees with u and the x-axis outside of this subregion (and within some fixed radius, say $1/4$) is closer to v than to u, and hence not in the Voronoi region of u. This follows clearly from the diagram in Figure 1. Letting w be such a point in the torus, then as in Figure 1, we have

$$d_{v,w}^2 = d_{u,w}^2 + d_{u,v}^2 - 2d_{u,w}d_{u,v}\cos a.$$

The angle a must be between 0 and 60 degrees and $\cos a$ is therefore greater than 1/2. Further, $d_{u,w} > d_{u,v}$. Hence $d_{v,w}^2 < d_{u,w}^2$.

We can conclude that if all six of the subregions around u contain another of the $n-1$ points, then the Voronoi region of u is contained within the circle of area c/n around u. This yields the lemma. ■

Using this lemma, we show the following.

Lemma 9 *When n points are placed independently and uniformly at random in the unit torus, the number of Voronoi regions of size at least c/n when $\ln n \geq c \geq 12$ exceeds $12ne^{-c/6}$ with probability $o(1/n^4)$.*

Proof: Let $Z_{i,j}$, $1 \leq i \leq n$ and $1 \leq j \leq 6$, be 1 if the jth subregion of area $c/6n$ around the ith point does not contain any of the other $n-1$ points and 0 otherwise. By Lemma 8, $Z = \sum_{i,j} Z_{i,j}$ is clearly an upper bound on the number of Voronoi regions of size at least c/n. Now

$$\mathbf{E}[Z] = 6n\left(1 - \frac{c}{6n}\right)^{n-1} < 6ne^{-c/6}.$$

Consider the n random points being placed in the circle one at a time. Let X_i be the location of the ith point, and $Y_i = \mathbf{E}[Z \mid X_1, X_2, \ldots, X_i]$. Then the sequence of Y_i is a Doob martingale with expectation $\mathbf{E}[Z]$. Unfortunately, it is not a useful Doob martingale in the context of using Azuma's inequality, since it does not obey a Lipschitz condition. That is, the introduction of a single point X_k can affect a significant number of the random variables $Z_{i,j}$.

To account for this difficulty, we introduce the following modification. Let $F = f(X_1, X_2, \ldots, X_n)$ be the total number of *empty-or-rare* subregions of area $c/6n$ around one of the X_i satisfying the following: either the subregion is empty, or for every X_j in the subregion, there exists $\ln^3 n$ points X_k with $k < i$ and X_j is in the circle of area c/n around X_k. Or, thinking of it another way, as we lay the points in order, each point can change the empty-or-rare status of at most $\ln^3 n$ subregions of other points, and they affect the subregions of the first $\ln^3 n$ points in the order of placement. The idea behind this change is that the rare regions are sufficiently rare that in almost all cases F and Z will be equal, since with high probability no point will ever affect the subregions of $\ln^3 n$ other points. In fact, as $F \geq Z$, we have

$$\mathbf{Pr}(Z \geq 12ne^{-c/6}) \leq \mathbf{Pr}(F \geq 12ne^{-c/6}),$$

so it suffices to bound the latter probability.

If we now let $Y_i = \mathbf{E}[F \mid X_1, X_2, \ldots, X_i]$, then the sequence of Y_i is a Doob martingale with expectation $\mathbf{E}[F]$. Now we can apply Azuma's inequality, since the placement of any point can affect only at most $\ln^3 n + 6$ subregions ($\ln^3 n$ subregions of other points, and the six subregions of the point itself). Hence this martingale does obey a Lipschitz condition with bound $\ln^3 n + 6$.

It remains to bound $\mathbf{E}[F]$. Clearly

$$\mathbf{E}[F] \leq \mathbf{E}[Z] + 6n\mathbf{Pr}(Z \neq F),$$

since when F and Z are not the same, the maximum possible value for F is 6n. To bound $\mathbf{Pr}(Z \neq F)$, note that for Z and F to differ, some point X_k must affect the value of some $Z_{i,j}$ for at least $\ln^3 n$

neighboring points. But this means there must be at least $\ln^3 n$ points in an area of size c/n around X_k, where $c \leq \ln n$. Using the standard Chernoff bound from [11], we have that the probability that this happens for any specific X_k is $n^{-O(\log\log n)}$, and hence $\mathbf{Pr}(Z \neq F) = n^{-O(\log\log n)}$. It follows easily that $\mathbf{E}[F] \leq 6n\mathrm{e}^{-c/6}$ as well for n sufficiently large. (That is, $\mathbf{E}[Z] = 6n\left(1 - \frac{c}{6n}\right)^{n-1}$, which differs from $6n\mathrm{e}^{-c/6}$ by more than $n^{-O(\log\log n)}$.) Azuma's inequality [11] yields

$$\mathbf{Pr}(F \geq 12n\mathrm{e}^{-c/6}) \leq \mathrm{e}^{-18n\mathrm{e}^{-c/3}/(\ln^3 n+6)}.$$

For c in the given range, this probability is certainly $o(1/n^4)$. ■

Once we have this bound, we can modify Lemma 6 and Theorem 1 to this setting as well. In the full version we explain the changes that need to be made. Minor changes involve modifying the constants. A slightly more subtle technical change is that the martingale bound only allows us to make statements bounding with high probability the length of the largest n^γ regions for some $\gamma < 1$. We must therefore be a bit more careful at the tail end of the argument of Theorem 1 once we get down to this few regions; however, since each region's area can be bounded by $O(\log n)$ with high probability, this causes no difficulty.

The key point is the exponential tail bound: we have that the number of regions of area c/n is bounded by $c_1 n\mathrm{e}^{-c/c_2}$ for appropriate constants c_1 and c_2 with high probability, and this is really all we need for Theorem 1 to hold. Hence Theorem 1 applies more generally to any two-choice scenario where we can determine such a tail bound. For example, the ideas of Lemmas 8 and 9 can be generalized to obtain similar bounds for higher constant dimension.

4 Experimental Results

Because the theoretical analysis introduces significant $O(1)$ constants, the power of two choices is best tested with experiments. Table 1 shows experiments for the case of points placed on random arcs, where in the case where two arcs have the same load, the tie is broken randomly. (The results are based on just 1000 trials.) It is clear that the maximum load increases steadily when just one choice is used, and the $O(\log\log n)$ behavior of using two (or more) choices is readily apparent. Similar results hold in the two-dimensional setting of the torus, as shown in Table 2.

Of course, breaking ties more carefully can yield better results, a concept first elucidated by Vöcking [15]. Our theoretical results are based on bounding the possible length of the arcs (or areas of the Voronoi regions) with large load. A natural approach is therefore to break ties by increasing the load on the smaller arc (or region). In Table 3 we examine the effect of this tie-breaking scheme, again for the case of points placed on random arcs. It does indeed slightly improve the maximum load, performing even slightly better than Vöcking's scheme. Clearly an interesting open problem would be to determine the exact performance of this variant, or at least to prove whether or not it is provably better asymptotically than breaking ties randomly.

n	$d = 1$		$d = 2$	$d = 3$	$d = 4$
2^8	**5** ...... 1.1% **6** ...... 12.3% **7** ...... 23.6% **8** ...... 23.9% **9** ...... 18.8% **10** ...... 9.6% **11** ...... 5.7% **12** 2.1%	**13** ...... 1.7% **14** ...... 0.4% **15** ...... 0.2% **16** ...... 0.4% **17** ...... 0.1% **18** ...... 0.1% **19** 0.1%	**3** ...... 26.8% **4** ...... 70.0% **5** 3.2%	**2** ...... 0.1% **3** ...... 97.9% **4** 2.0%	**2** ...... 13.1% **3** 86.9%
2^{12}	**9** ...... 0.9% **10** ...... 11.7% **11** ...... 23.8% **12** ...... 23.0% **13** ...... 18.9% **14** ...... 10.2% **15** ...... 5.3% **16** 3.0%	**17** ...... 1.3% **18** ...... 0.6% **19** ...... 0.7% **20** ...... 0.4% **21** ...... 0.1% **22** ...... 0.1% **24** 0.1%	**4** ...... 88.1% **5** ...... 11.8% **6** 0.1%	**3** ...... 89.6% **4** 10.4%	**3** 100.0%
2^{16}	**13** ...... 1.1% **14** ...... 12.6% **15** ...... 24.4% **16** ...... 22.0% **17** ...... 16.6% **18** ...... 11.2% **19** ...... 6.2% **20** 2.5%	**21** ...... 1.8% **22** ...... 0.6% **23** ...... 0.4% **24** ...... 0.1% **25** ...... 0.3% **26** ...... 0.1% **32** 0.1%	**4** ...... 19.6% **5** 80.4%	**3** ...... 21.0% **4** 79.0%	**3** 100.0%
2^{20}	**17** ...... 2.1% **18** ...... 11.4% **19** ...... 22.7% **20** ...... 21.0% **21** ...... 20.4% **22** ...... 10.3% **23** 6.3%	**24** ...... 2.3% **25** ...... 1.5% **26** ...... 1.0% **27** ...... 0.8% **28** ...... 0.1% **29** 0.1%	**5** ...... 99.9% **6** 0.1%	**4** 100.0%	**3** ...... 99.1% **4** 0.9%
2^{24}	**21** ...... 2.1% **22** ...... 9.7% **23** ...... 23.8% **24** ...... 23.8% **25** ...... 17.0% **26** ...... 10.9% **27** 5.6%	**28** ...... 3.3% **29** ...... 2.3% **30** ...... 0.8% **31** ...... 0.3% **32** ...... 0.2% **34** ...... 0.1% **35** 0.1%	**5** ...... 99.4% **6** 0.6%	**4** 100.0%	**3** ...... 86.5% **4** 13.5%

Table 1: Experimental maximum load with random arcs ($m = n$)

n	$d=1$	$d=2$	$d=3$	$d=4$
2^8	**4** ...... 4.0% **8** ...... 3.9% **5** ...... 38.4% **9** ...... 1.4% **6** ...... 35.5% **10** ...... 0.4% **7** ...... 16.3% **11** 0.1%	**2** ...... 0.2% **3** ...... 95.6% **4** 4.2%	**2** ...... 45.0% **3** 55.0%	**2** ...... 92.2% **3** 7.8%
2^{12}	**6** ...... 2.0% **10** ...... 5.8% **7** ...... 29.7% **11** ...... 1.5% **8** ...... 40.5% **12** ...... 0.2% **9** ...... 20.2% **13** 0.1%	**3** ...... 57.1% **4** 42.9%	**3** 100.0%	**2** ...... 31.9% **3** 68.1%
2^{16}	**8** ...... 0.7% **12** ...... 7.4% **9** ...... 26.9% **13** ...... 1.7% **10** ...... 44.1% **14** ...... 0.3% **11** ...... 18.8% **15** 0.1%	**4** 100.0%	**3** ...... 99.9% **4** 0.1%	**3** 100.0%
2^{20}	**10** ...... 0.9% **14** ...... 6.5% **11** ...... 22.0% **15** ...... 1.8% **12** ...... 45.7% **16** ...... 0.3% **13** 22.8%	**4** ...... 99.8% **5** 0.2%	**3** ...... 99.6% **4** 0.4%	**3** 100.0%

Table 2: Experimental maximum load with random torus polygons ($m = n$)

5 Conclusion

At the heart of this paper is an interesting generalization of the two-choice paradigm to geometric settings where the probability of choosing a bin is non-uniform. In distributed computing settings, such scenarios naturally arise when using nearest-neighbor methods in conjunction with hashing for random placement. This is exactly the setting in a Chord system without the use of virtual servers. Both through theoretical analysis and via experiments, we have demonstrated that the benefits of the two-choice paradigm extend to this setting.

It would be interesting to consider further related theoretical and practical questions. On the theoretical side, it is interesting to ask how much non-uniformity among bins the two-choice paradigm can stand. Perhaps one can make a general statement about what minimal properties a metric space requires for the two choice paradigm to apply. At the boundary of the theoretical and the practical, it would be an improvement if the theory could be used to accurately predict the resulting load distribution. In the case of uniform bin sizes, this can be done quite well using methods based on differential equations pioneered by Mitzenmacher [9]. While not as accurate as differential equaitons, the witness tree approach, as demonstrated by Vöcking [15], gives a somewhat tighter analysis than the original argument of Azar et al. It is not clear whether either of these methods can be made to apply to this setting, but perhaps some other argument can provide better results. On the practical side, while we believe the two-choice paradigm will prove useful for Chord-like networks, there is work to be done considering how to apply it while maintaining reliability and other useful features of these systems [3].

n	arc-larger	arc-random	arc-left	arc-smaller
2^8	**3** ...... 8.5% **4** ...... 82.8% **5** ...... 8.6% **6** 0.1%	**3** ...... 26.8% **4** ...... 70.0% **5** 3.2%	**3** ...... 57.3% **4** ...... 42.5% **5** 0.2%	**3** ...... 72.4% **4** 27.6%
2^{12}	**4** ...... 39.7% **5** ...... 60.2% **6** 0.1%	**4** ...... 88.1% **5** ...... 11.8% **6** 0.1%	**4** ...... 99.9% **5** 0.1%	**3** ...... 1.7% **4** ...... 97.9% **5** 0.4%
2^{16}	**5** ...... 99.6% **6** 0.4%	**4** ...... 19.6% **5** 80.4%	**4** ...... 96.7% **5** 3.3%	**4** ...... 99.0% **5** 1.0%
2^{20}	**5** ...... 93.9% **6** 6.1%	**5** ...... 99.9% **6** 0.1%	**4** ...... 63.9% **5** 36.1%	**4** ...... 88.8% **5** 11.2%
2^{24}	**5** ...... 37.4% **6** 62.6%	**5** ...... 99.4% **6** 0.6%	**5** 100.0%	**4** ...... 10.5% **5** 89.5%

Table 3: Experimental maximum load varying strategies for random arcs with $d = 2$ ($m = n$)

References

[1] Azar, Y., Broder, A., Karlin, A., and Upfal, E. Balanced allocations. *SIAM Journal of Computing 29*, 1 (1999), 180–200.

[2] Berenbrink, P., Czumaj, A., Steger, A. and Vöcking, B. Balanced allocations: the heavily loaded case. In *Proceedings of the 32nd ACM Symposium on Theory of Computing*, pp. 745-754, 2000.

[3] Byers, J., Considine, J., and Mitzenmacher, M. Simple Load Balancing for Distributed Hash Tables. BU Computer Science Technical Report 2002-29, November 2002. To appear in IPTPS 2003.

[4] Dabek, F., Kaashoek, M. F., Karger, D., Morris, R., and Stoica, I. Wide-area cooperative storage with CFS. In *Proceedings of the 18th ACM Symposium on Operating Systems Principles (SOSP '01)* (Chateau Lake Louise, Banff, Canada, Oct. 2001).

[5] Dubhashi, D.P. and Ranjan, D. Balls and bins: a case study in negative dependence. *Random Structures and Algorithms*, 13(2), pp. 99-124, 1998.

[6] Miles, R. E. A synopsis of 'Poisson flats in Euclidean spaces'. In: E. F. Harding and D. G. Kendall (eds), *Stochastic Geometry*. New York, John Wiley, pp. 202-227, 1974.

[7] Moller, J. Random tesselations in R^d. *Advances in Applied Probability*, 21, pp. 37-73, 1989.

[8] Karger, D. R., Lehman, E., Leighton, F. T., Panigrahy, R., Levine, M. S., and Lewin, D. Consistent hashing and random trees: Distributed caching protocols for relieving hot spots on the world wide web. In *ACM Symposium on Theory of Computing* (May 1997), pp. 654–663.

[9] Mitzenmacher, M. The power of two choices in randomized load balancing. Ph.D. thesis, U.C. Berkeley, 1996.

[10] Mitzenmacher, M., Richa, A., and Sitaraman, R. *The Power of Two Choices: A Survey of Techniques and Results.* Kluwer Academic Publishers, Norwell, MA, 2001, pp. 255–312. edited by P. Pardalos, S. Rajasekaran, J. Reif, and J. Rolim.

[11] Motwani, R. and Raghavan, P. Randomized Algorithms. Cambridge University Press, 1995.

[12] Ratnasamy, S., Francis, P., Handley, M., Karp, R., and Shenker, S. A scalable content addressable network. In *ACM SIGCOMM* (2001), pp. 161–172.

[13] Rowstron, A., and Druschel, P. Pastry: Scalable, distributed object location and routing for large-scale peer-to-peer systems. In *Proceedings of Middleware 2001* (2001).

[14] Stoica, I., Morris, R., Karger, D., Kaashoek, M. F., and Balakrishnan, H. Chord: A scalable peer-to-peer lookup service for internet applications. In *ACM SIGCOMM* (2001), pp. 149–160.

[15] Vöcking, B. How asymmetry helps load balancing. In *Proceedings of the* 40^{th} *IEEE-FOCS* (1999), pp. 131–140.

[16] Zhao, B. Y., Kubiatowicz, J. D., and Joseph, A. D. Tapestry: An infrastructure for fault-tolerant wide-area location and routing. Tech. Rep. UCB/CSD-01-1141, UC Berkeley, Apr. 2001.

6 Appendix

We now prove the following technical claim, used in the proof of Theorem 1.

Claim 10 *We have* $i^* = \frac{\log\log n}{\log d} + O(1)$.

Proof: To prove the claim, first note that

$$\beta_{i+1} \geq \frac{\beta_i^d}{n^{d-1}},$$

so a simple induction yields that

$$\ln \frac{n}{\beta_{i+256}} \leq d^i \ln \frac{n}{\beta_{256}}.$$

We now claim that, using the above, a simple induction shows that for $k \geq 2$

$$\beta_{k+256} \leq n \left(\frac{\beta_{256}}{n}\right)^{d^k} 2^{d^k+1+2\sum_{j=1}^{k-1} d^j} \left(\ln \frac{n}{\beta_{256}}\right)^{\sum_{j=1}^{k} d^j} d^{\sum_{j=1}^{k-1}(k-j)d^j}. \tag{2}$$

To see this, note that

$$\beta_{257} = 2n\left(2\frac{\beta_{256}}{n} \ln \frac{n}{\beta_{256}}\right)^d = n\left(\frac{\beta_{256}}{n}\right)^d 2^{d+1}\left(\ln\frac{n}{\beta_{256}}\right)^d,$$

and

$$\begin{aligned}
\beta_{258} &= 2n\left(2\frac{\beta_{257}}{n}\ln\frac{n}{\beta_{257}}\right)^d \\
&\leq 2n\left(2\frac{\beta_{257}}{n}d\ln\frac{n}{\beta_{256}}\right)^d \\
&= n\left(\frac{\beta_{256}}{n}\right)^{d^2} 2^{d^2+2d+1}\left(\ln\frac{n}{\beta_{256}}\right)^{d^2+d} d^d,
\end{aligned}$$

giving the base case. The induction is now clear using

$$\begin{aligned}
\beta_{k+256} &= 2n\left(2\frac{\beta_{(k-1)+256}}{n}\ln\frac{n}{\beta_{(k-1)+256}}\right)^d \\
&\leq 2n\left(2\frac{\beta_{(k-1)+256}}{n}d^{k-1}\ln\frac{n}{\beta_{256}}\right)^d
\end{aligned}$$

and plugging in equation (2). Bounding the terms of (2) with more pleasant expressions, we have

$$\beta_{k+256} \leq n\left(\frac{\beta_{256}}{n}\right)^{d^k} 8^{d^k}\left(\ln\frac{n}{\beta_{256}}\right)^{2d^k} d^{4d^{k-1}}.$$

It follows since $\beta_{256} = n/256$ that

$$\beta_{k+64} \leq n\left(\frac{8d^{4/d}\ln 256}{256}\right)^{d^k} \leq c^{d^k}$$

for some constant $c < 1$ (for any integer value of $d \geq 2$). That i^* is $\frac{\log\log n}{\log d} + O(1)$ now follows. ∎