ECE 559BH Distributed Network Algorithms **SPRING 2006**
PROBLEM SET 1 **Due 2/2/6**

Balls in bins, Chord peer-to-peer lookup service , and BitTorrent file distribution system

Reading: Section 3.6 of Motwani and Raghavan, *Randomized Algorithms* and Chapter 5 of Mitzenmacher and Upfal, *Probability and Computing* cover the coupon collectors problem and the Poisson heuristic and associated bounds. See course website for reading on Chord and BitTorrent.

Problems to be handed in:

1. Splitting a Poisson random variable
(This problem concerns the key facts underlying the Poisson heuristic and associated bounds.) (a) Let $\lambda > 0$. Suppose that for each integer $n \geq \lambda$, Y_n is a random variable with the binomial distribution with parameters n and $p = \lambda/n$. Show that $\lim_{n\to\infty} P[Y_n = k] = \frac{e^{-\lambda}\lambda^k}{k!}$. This result reflects the fact that a binomial(n,p) random with large n and small p has approximately a Poisson distribution.
(b) Consider a random number X of objects, where X is a Poisson random variable with mean $\lambda > 0$. A sample of X objects is drawn, with each object being one of n types. Let $(p_1, \ldots, p_n)$ be a probability vector, and suppose that for each k, the k^{th} object is type i with probability p_i, independently of the types of the other objects. Let X_i denote the total number of objects drawn of type i. Show that $X_1, \ldots, X_n$ are mutually independent, and that X_i has the Poisson distribution with mean λp_i. (Hint: try $n = 2$ first.)

2. The near median of the Poisson distribution
Suppose Z is a Poisson random variable with mean m, where m is an integer with $m \geq 1$, and let $p_i = P\{Z = i\}$. Show that $p_{m+i} \geq p_{\mu-i-1}$ for $0 \leq i \leq m-1$, and conclude that $P\{Z \geq m\} \geq \frac{1}{2}$. (Note, it can in fact be shown that if Z is the sum of independent 0-1 valued random variables with an integer mean m, then m is also the median of the distribution. (K. Jogdeo and S. Samuels, "Monotone convergence of binomial probabilities and a generalization of Ramanujan's equation," *Annals of Math. Stat.*, vol. 39 (1968), pp. 1191-1195.)

3. Large deviations for the coupon collectors problem
As shown in class using the Poisson heuristic and associated bounds, if X is the number of coupons observed before obtaining one of each of n types, $\lim_{n\to\infty} P\{X > n \ln n + cn\} = 1 - e^{-e^{-c}}$ for each c fixed. This shows that deviations of X on the order of n are typical. Using the Poission method and the bound of the previous problem, provide upper bounds on $P\{X < (1-\epsilon)n \ln n\}$ and $P\{X > (1+\epsilon)n \ln n\}$ which converge to zero quickly as n grows, for $\epsilon > 0$ fixed.

4. Chord lookups
(a) Explain why the mean number of nodes contacted for a lookup is about $0.5 \log_2 N$.
(b) Suppose that each node in a chord runs the stabilize routine (see Figure 7 of the paper) at least once every 10 seconds, and that execution of stabilize is essentially instantaneous. Suppose a and b are nodes on the chord such that, initially, $b = a.successor$. Then suppose L new nodes are added to the chord, such that all L nodes are in the interval (a, b). Determine the maximum amount of time that would be needed until the predecessor and successor variables are correct for all the nodes.

5. BitTorrent choking algorithms (S. Shakkottai)
(a) Explain how the dynamics of BitTorrent (as described in Bram Cohen's paper) tends to cause peers to typically download from other peers who have roughly the same bandwidth connections to the Internet (assume the upload bandwidth of any given peer is the same as its download bandwidth). Explain an advantage of this.
(b) The optimistic unchoking mechanism used by peers in BitTorrent offers a peer the possibility of downloading without ever uploading. According to the model of D. Qui and R. Srikant, the downloading rate of such a free-riding peer would be about one-fifth the average bandwidth of its peers. Would a peer have incentive to free-ride if all peers in the network have the same bandwidth? How about if there is a wide range of bandwidths among the peers of the network?