Maximum Level in a Skip List

Say we construct a skip list by inserting n elements in arbitrary order, choosing the level of each element using the *randomLevel* function described in the paper (first setting *MaxLevel* to infinity, so that *newLevel* is always output).

The probability that a particular element reaches level at least k is p^k, and the probability that *any* of the n elements reach level at least k is at most np^k.

Let M be the maximum level of any element in the skip list. We want to bound $\mathrm{E}[M]$ given that $\Pr[M \geq k] \leq np^k$. Using our usual technique of partitioning the possible experiments up using some group of mutually exclusive events, we have:

$$\mathrm{E}[M] = \sum_{k=0}^{\infty} kPr[M = k] \leq \sum_{k=0}^{\infty} kPr[M \geq k]$$

Unfortunately using this in a totally straighforward way leads to a ridiculously high upper bound. We get:

$$\mathrm{E}[M] \leq \sum_{k=0}^{\infty} knp^k = n \sum_{k=0}^{\infty} kp^k$$

This last sum seems simple enough to look up, and we find that it is $p/(1-p)^2$, for $0 \leq p < 1$, a constant. So we have shown that $E[M] = O(n)$, not very helpful.

What went wrong? The low terms in the sum are huge over-estimates. For instance, if $p = 1/2$ we have

$$Pr[M \geq 1] \leq n/2$$

which is true, but not a very good upper bound for a probability. When does it start getting to be a useful bound? Around when $k = \lg n$:

$$Pr[M \geq \lg n] \leq n/n = 1$$

To get a tighter bound, we break the sum into two parts (this is a handy technique! especially for Homework problem 5!). We'll choose a number L which will be the boundary between small values of k and large values of k. Then we'll break the sum at L:

$$\mathrm{E}[M] \leq \sum_{k=0}^{L-1} kPr[M = k] + \sum_{k=L}^{\infty} kPr[M = k]$$

On the part with large k, we'll use the upper bound, and we'll find some other way to handle the small k.

So how do we choose L? When is k large enough? We'll choose L so that

$$knp^k = O(1/k^2), \ \forall k \geq L$$

Why $1/k^2$? Because another of the essential sums one ought to know is that

$$\sum_{i=0}^{\infty} 1/i^2 \leq 2$$

so that

$$\sum_{k=L}^{\infty} knp^k \leq \sum_{k=L}^{\infty} O(1/k^2) = O(1)$$

So what exactly is L? We want

$$Lnp^L \leq 1/L^2$$

for large enough values of n; so we want to choose L, as a function of n and p, so that

$$L^3 p^L = o(1/n)$$

Solving directly for L is difficult. Instead, we just plug in some values and find one that works, preferably the smallest one possible that works. A good choice ends up being

$$L = 2\lg_{1/p} n$$

(plug it in and check that $Lnp^L = o(1/L^2)$!).

Now we just need to figure out what to do with the small-k terms. Fortunately there are very few of these ($O(\lg n)$, because of our choice of L) so we can make some generous over-estimates;

$$\sum_{k=0}^{L-1} kPr[M = k] \leq \sum_{k=0}^{L-1} LPr[M = k] = L\sum_{k=0}^{L-1} Pr[M = k] = LPr[M < L] \leq L$$

So we end up with

$$E[M] = \sum_{k=0}^{L-1} kPr[M = k] + \sum_{k=L}^{\infty} kPr[M = k] \leq L + O = O(\lg n)$$