# Lesson 2. Notes on Skip lists

M. Pellegrini*

January 19, 2004

## 1 Disclaimer

*These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

## 2 Introduction

This lecture is a mix of [MR95, Chapter 1] and [Mul93, Chapter 1].

In the slides we have:

1. Proof that the height is $O(\log n)$ with high probability.
2. Proof of expected linear storage.
3. Quasi-Proof of expected $O(\log n)$ search time (actually $O(1)$ expected per level.
4. Proof that insertion/deletion are done in search time plus expected $O(1)$ time.

## 3 Proper analysis of skip lists search time form [Smi94]

Given the skip list and a query $x$ let $C_i$ be the number of elements in the i-th list to which $x$ is compared during the search algorithm. Let $h$ be the height of the data structure. The search cost is proportional to $\sum_{i=1}^{h}(1 + C_i)$.

Unfortunately we cannot conclude that the expected cost is proportional to $\sum_{i=1}^{E[h]}(1+E[C_i])$ because the linearity of expectation is proved only for a summation with a fixed

*Istituto di Informatica e Telematica del CNR, Via G. Moruzzi 1, 56100-Pisa (Italy). pellegrini@iit.cnr.it

number of random variables. Therefore the proof has to be more complex. We fix a level $A$ and we analyze separately the expectation of the costs on levels $1,..,A$ and the expectation for levels $A+1,..,h$. Afterwards we will decide a convenient value for $A$.

Case 1. Consider levels from $A+1,..,h$. the cost of searching in those lists is for sure bounded by the total size of them: $M(A)=\sum_{i=A+1}^{h}|L_i|$. Note that the lists $L_i$ for $i=A+1,..h$ form a skip list for the set $S_{A+1}$. We use the notion of conditional expectation:

$$E[M(A)]=\sum_{k=0}^{n}E[M(A)||S_{A+1}|=k]Pr(|S_{A+1}|=k)$$

Now $E[M(A)||S_{A+1}|=k]$ is just the expected size of a skip list for a $k$-elements set, which is $O(k)$. Now we estimate $Pr(|S_{A+1}|=k)$. There are $\binom{n}{k}$ k-subsets of $n$ elements. The probability of a given element to reach level $A+1$ is $1/2^A$, the probability of not reaching it is $1-(1/2^A)$. Thus

$$Pr(|S_{A+1}|=k)=\binom{n}{k}(1/2^A)^k(1-(1/2^A))^{n-k}$$

Set $1/2^A=p$, the expected cost is order of:

$$\sum_{k=0}^{n}k\binom{n}{k}p^k(1-p)^{n-k}$$

Use the identity

$$\binom{n}{k}=\frac{n}{k}\binom{n-1}{k-1}$$

collect $p$, rescale $k'=k-1$, use the Newton expansion of $(a+b)^x$.

$$\sum_{k=1}^{n}n\binom{n-1}{k-1}p^k(1-p)^{n-k}=$$

$$np\sum_{k=1}^{n}\binom{n-1}{k-1}p^{k-1}(1-p)^{n-k}=$$

$$np \sum_{k'=0}^{n-1} \binom{n-1}{k'} p^{k'} (1-p)^{n-1-k'} =$$

$$np(p + (1-p))^{n-1} = np.$$

Thus we have proved a bound $O(n/2^A)$ for the levels $A..h$.

Case 2. levels 1,..,A.

Now we can use the formula:

$$E[\sum_{i=1}^{A} (1 + C_i)] = \sum_{i=1}^{A} (1 + E[C_i]) = O(1)A = O(A)$$

So the expected cost is bounded by $O(n/2^A + A)$ taking $A = \lceil \log n \rceil$ we have a search time $O(\log n)$.

# 4 Notes

In [Smi94] one finds also concentration bounds for skip list performance obtained using the Chernoff bounds technique.

# 5 Exercises

(1) Repeat the analysis of skip lists for a coin with bias $p$, with $0 < p < 1$.

# References


[MR95] R. Motwani and P. Raghavan. *Randomized Algorithms.* Cambridge University Press, 1995.

[Mul93] K. Mulmuley. *Computational Geometry, an Introduction through Randomized Algorithms.* Prentice Hall, 1993.

[Smi94] M. Smid. Selected topics in data structures. Technical Report MPI-I-94-155, Max Plank Institute fuer Informatik, December 1994.