INFORMATIQUE THÉORIQUE ET APPLICATIONS

L. DEVROYE
P. KRUSZEWSKI

On the Horton-Strahler number for random tries

Informatique théorique et applications, tome 30, n° 5 (1996), p. 443-456

<http://www.numdam.org/item?id=ITA_1996__30_5_443_0>

Informatique théorique et Applications/Theoretical Informatics and Applications
(vol. 30, n° 5, 1996, pp. 443-456)

ON THE HORTON-STRAHLER NUMBER FOR RANDOM TRIES (*)

by L. DEVROYE ([1]) and P. KRUSZEWSKI ([2])

Communicated by A. ARNOLD

Abstract. – *We consider random tries constructed from n i.i.d. sequences of independent Bernoulli (p) random variables, $0 < p < 1$. We study the Horton-Strahler number H_n, and show that*

$$\frac{H_n}{\log n} \to \frac{1}{\log \frac{1}{\min(p,1-p)}}$$

in probability as $n \to \infty$.

Keywords: Horton-Strahler number, trie, probabilistic analysis, data structures, random trees.

Résumé. – *On étudie des arbres aléatoires du type « trie » construits à partir de n suites indépendantes de variables aléatoires Bernoulli (p) où $0 < p < 1$. On prouve que*

$$\frac{H_n}{\log n} \to \frac{1}{\log \frac{1}{\min(p,1-p)}}$$

en probabilité, où H_n est le nombre de Horton-Strahler.

INTRODUCTION

In 1960, Fredkin [9] coined the term *trie* for an efficient data structure to store and re*trieve* strings. These were further developed and modified by Knuth [4], Larson [16], Fagin, Nievergelt, Pippenger and Strong [6], Litwin [17], Aho, Hopcroft and Ullman [1] and others. The tries considered

(*) Received September 1995.

([1]) School of Computer Science, McGill University, 3480 University Street, Montreal, Canada H3A2A7, Research supported by NSERC Grant A3456 and FCAR Grant 90-ER-0291. Email: luc@cs.mcgill.ca

([2]) School of Computer Science, McGill University, 3480 University Street, Montreal, Canada H3A2A7, Research supported by a 1967 NSERC Postgraduate Scholarship. Email: kruz@cs.mcgill.ca

Informatique théorique et Applications/Theoretical Informatics and Applications
0988-3754/96/05/$ 7.00/

here are constructed from n independent infinite binary strings $X_1, \ldots, X_n$. Each string defines an infinite path in a binary tree: a 0 forces a move to the left, and a 1 forces a move to the right. An *infinite p-trie* is a random binary tree obtained by highlighting n infinite paths (from the root down). These paths are independent and are described by independent, identically distributed (i.i.d.) sequences of Bernoulli (p) random variables, $0 < p < 1$. For example, Figure 1 shows an infinite p-trie built from the infinite strings $01001\ldots$, $01011\ldots$, $10011\ldots$, $10100\ldots$ and $11100\ldots$. The tree is now pruned so that it has just n leaves at the n representative nodes (e.g., see *Fig.* 2). That is, the *finite p-trie* is the infinite p-trie maximally trimmed so that each of the n infinite paths is finite and visits at least one node not visited by any other path (that node is necessarily a leaf of the future p-trie). Observe that no representative node is allowed to be an ancestor of any other representative node. This implies that every internal (non-leaf) node has at least two leaves in its collection of descendants.

Originally used to classify river systems by Horton [11] and Strahler [24], the Horton-Strahler number has also been applied to binary trees. Let u be a

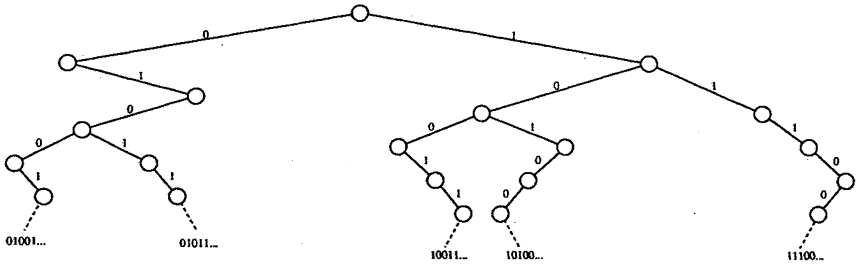

Figure 1. – An infinite p-trie.

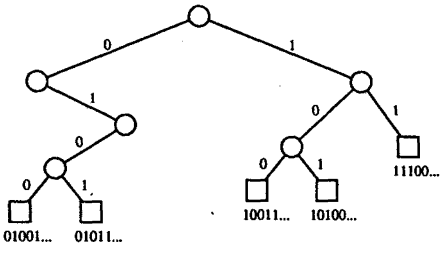

Figure 2. – The p-trie is a trimmed-down version of the infinite p-trie in which the strings are associated with the leaves.

node in a binary tree. Let $|u|$ be the number of nodes in the subtree rooted at u (with u included) and let the Horton-Strahler number S_u be defined by

$$S = \begin{cases} 0 & \text{if } |u| = 0, \\ \max\{S_v, S_w\} + I_{[S_v = S_w]} & \text{if } |u| \geq 1 \text{ and } u \text{ has (possibility-nonexistent) children } v \text{ and } w, \end{cases}$$

where I is the indicator function. Note that leaves u have $S_u = 1$, and that internal nodes u with one proper child v have $S_u = S_v$.

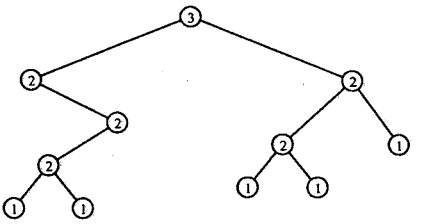

Figure 3. – A binary tree with Horton-Strahler labelling.

In general, let H_n be the Horton-Strahler number of the root of a binary tree with n nodes. For a chain-shaped tree, $H_n = 1$. For a complete tree with k full levels and $2^k - 1$ nodes, we have $H_n = k$. A little thought shows that $H_n \leq \log_2 n + 1$.

The Horton-Strahler number arises in computer science because of its relationship to expression evaluation. In a computer, an arithmetic expression is evaluated by micro-operations using registers. To facilitate this process, the expression is stored as an expression tree with the operators in the internal nodes and the operands in the external nodes. The arithmetic expression is evaluated by traversing the corresponding tree. In 1958, Ershov [5] showed that by always traversing the child node with the lower Horton-Strahler number first, the corresponding register use is minimal (note however that this does not minimize time). Furthermore, the minimum number of registers required to evaluate an expression tree with root u is exactly $S_u + 1$. As expression evaluation is a special type of postorder traversal, the same paradigm shows that the minimum stack size required for a postorder traversal of a binary tree is $S_u + 1$ (e.g., *see* Françon [8]). In fact, the Horton-Strahler number occurs in almost every field involving some kind of natural branching

pattern. More recently, the Horton-Strahler number has been used to draw trees by Viennot, Eyrolles, Janey and Arquès [29] and Kruszewski [15]. Viennot [28] provides a thorough overview. *See* also Vauchaussade de Chaumont [26] and Vauchaussade de Chaumont and Viennot [27].

The properties of the Horton-Strahler number have only been studies for one model of random binary tree, equiprobable binary trees (EBT). These are random binary trees with n nodes drawn uniformly and at random from all possible rooted binary trees with n nodes. Let H_n be the Horton-Strahler number of a random EBT with n nodes so that $\mathbf{E}H_n$ and $\mathbf{Var}\{H_n\}$ are the corresponding expected value and variance. It is well-known (*see*, e.g., Flajolet, Raoult and Vuillemin [7], Kemp [13], Meir and Moon [18], Meir, Moon and Pounder [19], Moon [20], Devroye and Kruszewski [4], and Prodinger [23]) that

$$\mathbf{E}H_n \sim \log_4 n \qquad \text{and} \qquad \mathbf{Var}\{H_n\} = O(1).$$

Viennot *et al.* [29] introduced the notion of corresponding ramification matrix. Penaud [21] proved their conjecture on the structure of the ramification matrix for EBTs. Viennot *et al.* [29] experimentally studied the ramification matrix for random binary search trees. Vannimenus and Viennot [25] experimentally studied the ramification matrix for "injection patterns".

In tree-drawing applications, one needs a family of trees with one or more parameters so that the resulting trees cover a sufficiently wide range of shapes. One such family is the family of tries with parameter p. As H_n varies with p, the parameter p may be used to control the "bushiness" and elongation of the drawn trees. For example, Arquès *et al.* [2] visualized tries as botanical trees. It would be desirable to have simple two- and multi-parameter families as well, for added flexibility. These may be obtained by considering Markovian tries, in which the i.i.d. Bernoulli sequences are replaced by Markovian sequences of random bits (*see*, e.g., Jacquet and Szpankowski [12]). The study of the Horton-Strahler number for this moel is not attempted here.

We first define two tree metrics, the Balance number and the Fill level, which serve as deterministic upper and lower bounds for the Horton-Strahler number. We then derive the upper and lower bounds respectively of these two metrics and show that they converge to the same value, thereby squeezing Horton-Strahler number between them.

THE BALANCE NUMBER

We first define an infinite trie T^* as the infinite complete binary tree. A position of a node in T^* is addressed by two integers, (i, l), where l is the level number ($l \geq 0$), and $0 \leq i \leq 2^l - 1$ is an integer indicating the node at level l. For example, the root is at level 0, so $i = l = 0$ for the root. The integer i when expanded into l bits describes the path from the root to the node (0 forces a left turn, 1 forces a right turn). Let $|i|_l$ denote the number of one bits in the last l bits of i.

If we take an i.i.d. sequence of Bernoulli (p) random variables, say $Z_1, Z_2, Z_3, \ldots$ and write the bits backwards to form integers, then we obtain the integers

$$Z_1 + 2^1 Z_2 + 2^2 Z_2 + \cdots.$$

These are precisely the integers visited on the path from the root by our sequence. At level 0, we visit 0. At level 1, Z_1, at level 2, $Z_1 + 2^1 Z_2$, and so forth. When we refer to node (i, l), and $i \geq 2^l$, we are in fact referring to $(i \bmod 2^l, l)$. Therefore, we allow such references modulo 2^l.

The probability that a random i.i.d. sequence of Bernoulli (p) random variables carves out a path that reaches (i, l) is $q_{i,l} = p^{|i|_l}(1-p)^{l-|i|_l}$. We call this the probability of node (i, l). For every node (i, l) we record its cardinality $C_{i,l}$, the number of the n strings $X_1, \ldots, X_n$ that go through it, *i.e.*, those strings that have in their first l bits the integer i written backwards. If $|i|_l = k$, then $C_{i,l}$ is binomial $(n, p^k(1-p)^{l-k})$. The sibling of a node (i, l) is (i', l) where i' and i differ in the last bit only. We define the Balance number of (i, l) as

$$B_{i,l} = \sum_{j=1}^{l} I_{[1 \leq C_{i,j} \leq C_{i',j}]}$$

where (i, j) denotes $(i \bmod 2^j, j)$. The Balance number B_n of the p-trie is

$$B_n = \sup_{(i,l)} B_{i,l},$$

where the supremum is only over those nodes (i, l) that are in the p-trie. For example, Figure 4 shows our trie with the edges labelled by the indicator function $I_{[1 \leq C_{i,j} \leq C_{i',j}]}$ and the nodes labelled by Balance number.

We note that since nodes with no siblings have the same Balance numbers as their parents, the finite and infinite p-tries (and the corresponding Patricia

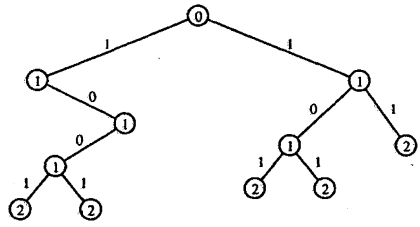

Figure 4. – The trie with Balance numbers.

tree – a Patricia tree is a trie in which all internal nodes with one child are removed and recursively replaced by that sole child) all have the same Balance number.

We now show the following upper bound on B_n.

THEOREM 1: *For* $0 < p < \frac{1}{2}$ *and* $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbf{P}\{B_n > (1+\varepsilon)\log_{\frac{1}{p}} n\} = 0.$$

Proof: The nodes are separated into three categories:

$$A = \{(i,l) \ : nq_{i,l} \geq n^{\varepsilon}\},$$
$$B = \{(i,l) \ : n^{-\varepsilon} < nq_{i,l} < n^{\varepsilon}\},$$
$$C = \{(i,l) \ : nq_{i,l} \leq n^{-\varepsilon}\}.$$

Let A_0 be the event that for all $(i,l) \in A$, $l \geq 1$, $C_{i,l} < C_{i',l}$ if and only if $q_{i,l} < q_{i',l}$. For $p < \frac{1}{2}$, we will see in Lemma 1 that, if A_0^c is the complement of A_0,

$$\mathbf{P}\{A_0^c\} \to 0.$$

On any path, the number of nodes that belong to B is not more than $2 + 2\varepsilon \log_{\frac{1}{1-p}} n$ (assuming still $p < \frac{1}{2}$, then paths of the form $(\ldots 000 \ldots)$ maximize the path length). Finally, let B^* be the subset of nodes in B with at least one child in C. We show in Lemma 2 that

$$\mathbf{P}\{\exists (i,l) \in B^* : C_{i,l} > M\} \to 0 \qquad \text{if} \quad M \geq 1 + \frac{1}{\varepsilon}. \tag{1}$$

Collecting all this, we note that for any (i,l), with probability tending to one,

$$B_{i,l} \leq M + \left(2 + 2\varepsilon \log_{\frac{1}{1-p}} n\right) + \left| \left\{ \begin{matrix} (m,j) \text{ on path from } (i,l) \text{ to root,} \\ (m,j) \in A,\ q_{mj} < q_{m'j} \end{matrix} \right\} \right| \tag{2}$$

As any path visits B^*, and every node of B^* has cardinality $\leq M$ with probability tending to one, the contribution to $B_{i,l}$ from all nodes below that node of B^* is $\leq M$. Observe that the last quantity of (2) is maximized by choosing i with binary expansion $(111\ldots)$.

Then we must have, for any $(m,j) \in A$ on the path to (i,l), $np^m \geq n^\varepsilon$, or $m \leq (1-\varepsilon)\log_{\frac{1}{p}} n$. Therefore, as we may take $M = 1 + \frac{1}{\varepsilon}$,

$$\begin{aligned} B_{i,l} &\leq 1 + \frac{1}{\varepsilon} + \left(2 + 2\varepsilon \frac{\log \frac{1}{p}}{\log \frac{1}{1-p}} \log_{\frac{1}{p}} n\right) + (1-\varepsilon)\log_{\frac{1}{p}} n \\ &\leq 3 + \frac{1}{\varepsilon} + \left(1 + 2\varepsilon \frac{\log \frac{1}{p}}{\log \frac{1}{1-p}}\right) \log_{\frac{1}{p}} n. \end{aligned}$$

Thus we will have shown that

$$\mathbf{P}\left\{ \sup_{(i,l)} B_{i,l} > \left(1 + 2\varepsilon \frac{\log \frac{1}{p}}{\log \frac{1}{1-p}}\right) \log_{\frac{1}{p}} n \right\} \to 0,$$

for all $\varepsilon > 0$. □

We are left with two technical lemmas.

LEMMA 1: $\mathbf{P}\{A_0^c\} \to 0$.

Proof: Take $(i,l) \in A$ and let (i^*, l^*) denote its parent (note: $l^* = l - 1$, $i^* = i \bmod 2^{l^*}$). Given C_{i^*,l^*}, we know that $C_{i,l}$ is binomial $(C_{i^*,l^*}, 1-p)$ or binomial (C_{i^*,l^*}, p) depending upon whether its is left or right child. Now, if $q_{il} < q_{i',l}$

$$\begin{aligned} [C_{i,l} \geq C_{i',l}] &= [C_{i,l} \geq C_{i^*,l^*} - C_{i,l}] \\ &= [C_{i,l} \geq \frac{1}{2} C_{i^*,l^*}] \\ &= \left[C_{i,l} - pC_{i^*,l^*} \geq \left(\frac{1}{2} - p\right) C_{i^*,l^*}\right]. \end{aligned}$$

Thus, by Hoeffding's inequality [10],

$$\mathbf{P}\{C_{i,l} \geq C_{i',l} | C_{i^*,l^*}\} \leq \exp\left\{-2\left(\frac{1}{2}-p\right)^2 C_{i^*,l^*}\right\}.$$

We argue similarly for $q_{i,l} > q_{i',l}$, $[C_{i,l} \leq C_{i',l}]$, and note that

$$\begin{aligned}
\mathbf{P}\left\{\begin{bmatrix} C_{i,l} \geq C_{i',l}, q_{i,l} < q_{i',l} \\ \text{or} \\ C_{i,l} \leq C_{i',l}, q_{i,l} > q_{i',l} \end{bmatrix}\right\} &\leq 2\mathbf{E}\{e^{-2(\frac{1}{2}-p)^2 C_{i^*,l^*}}\} \\
&\stackrel{\text{def}}{=} 2\mathbf{E}\{\delta^{C_{i^*,l^*}}\} \quad (\text{where } 0 < \delta < 1) \\
&= 2(1 - q_{i^*,l^*} + q_{i^*,l^*}\delta)^n \\
&\leq 2e^{-(1-\delta)q_{i^*,l^*}n} \\
&\leq 2e^{-(1-\delta)n^\varepsilon}
\end{aligned}$$

as $nq_{i^*,l^*} \geq n^\varepsilon$ because $(i,l) \in A$. Thus, by Boole's inequality,

$$\mathbf{P}\left\{\bigcup_{(i,l)\in A}\begin{bmatrix} C_{i,l} \geq C_{i',l}, q_{i,l} < q_{i',l} \\ \text{or} \\ C_{i,l} \leq C_{i',l}, q_{i,l} > q_{i',l} \end{bmatrix}\right\} \leq |A| 2e^{-(1-\delta)n^\varepsilon}. \tag{3}$$

Clearly, $|A|$ is not more than the number of leaves in the tree pruned to A times the height of A. But as the leaves are disjoint, their probabilities cannot sum to more than one, and each individual probability is at least $n^{-(1-\varepsilon)}$, the number is not more than $n^{1-\varepsilon}$. The height of A is not more than $1 + \log_{\frac{1}{1-p}} n$, by a trivial argument. Thus, (3) is not larger than

$$2\left(1 + \log_{\frac{1}{1-p}} n\right) n^{1-\varepsilon} e^{-(1-\delta)n^\varepsilon} \to 0. \quad \square$$

LEMMA 2.

$$\mathbf{P}\left\{\sup_{(i,l)\in B^*} C_{i,l} > M\right\} \to 0 \qquad for \quad M \geq 1 + \frac{1}{\varepsilon}.$$

Proof: First we count the number of nodes in B^*. Clearly, for any node in B^*, $nq_{i,l} > n^{-\varepsilon}$ and $nq_{i,l}p \leq n^{-\varepsilon}$ because one of its children must be in C. Let C^* be the collection of all the rightmost ("p") children of nodes in B^* (*i.e.*, all nodes in C^* have probability p times that of their parent in

B^*). Note that the nodes in C^* are disjoint, hence their probabilities sum to at most one. But for $(i,l) \in C^*$,

$$nq_{i,l} = nq_{i^*,l^*}p > n^{-\varepsilon}p,$$

or $q_{i,l} > p/n^{1+\varepsilon}$. Therefore, $|C^*|n^{-(1+\varepsilon)}p < 1$. Thus, $|B^*| < n^{1+\varepsilon}/p$. Fix $(i,l) \in B^*$. Recall that $q_{i,l} \leq 1/pn^{1+\varepsilon}$. Then

$$\begin{aligned}
\mathbf{P}\{C_{i,l} > M\} &\leq \sum_{j>M}^{n} \binom{n}{j} (q_{i,l})^j (1-q_{i,l})^{n-j} \\
&\leq \binom{n}{m} (q_{i,l})^m (1 + nq_{i,l} + (nq_{i,l})^2 + \ldots) \\
&\quad (\text{where } m = \lfloor M+1 \rfloor) \\
&\leq (nq_{i,l})^m \frac{1}{1-nq_{i,l}} \\
&\leq \frac{1}{(pn^\varepsilon)^m} \frac{1}{1-\frac{1}{pn^\varepsilon}} \\
&\leq \frac{2}{(pn^\varepsilon)^m}
\end{aligned}$$

for n large enough. Thus

$$\mathbf{P}\left\{ \sup_{(i,l)\in B^*} C_{i,l} > M \right\} \leq \frac{2|B^*|}{(pn^\varepsilon)^m} \leq \frac{2n^{1+\varepsilon}}{p(pn^\varepsilon)^m}$$

for all n large enough. This tends to zero if $\varepsilon m > 1+\varepsilon$. That is, if $m > 1 + 1/\varepsilon$. This holds if $M = 1 + 1/\varepsilon$. □

We can now derive the result in Theorem 1 for all $p \in (0,1)$.

COROLLARY 1: *For all* $\varepsilon > 0$, $0 < p < 1$,

$$\lim_{n\to\infty} \mathbf{P}\left\{ B_n > (1+\varepsilon) \log_{\frac{1}{\min(p,1-p)}} n \right\} = 0.$$

Proof: We note that for $p = 1/2$, the same proof works throughout, except for the following. From (2), regardless of whether Lemma 1 holds or not,

$$\begin{aligned}
B_{i,l} &\leq M + 2 + 2\varepsilon \log_2 n + (1-\varepsilon)\log_2 n \\
&\leq 1 + \frac{1}{\varepsilon} + (1+\varepsilon)\log_2 n.
\end{aligned}$$

So, we need not bother with (2) nor an extension of Lemma 1. In the proof of Lemma 2, the fact that $p < 1/2$ was not used. We thus see that for all $\varepsilon > 0$, $0 < p < 1$,

$$\lim_{n\to\infty} \mathbf{P}\left\{B_n > (1+\varepsilon)\log_{\frac{1}{\min(p,1-p)}} n\right\} = 0. \quad \square$$

THE FILL LEVEL

The Fill level or saturation level of a binary tree is the deepest level l in the tree such that all possible 2^l nodes at that level exist. For example, the trie of Figure 2 has Fill level 2. In 1992, Devroye [3] showed that for random Patricia trees constructed from n i.i.d. sequences of independent equiprobable bits and Fill level F_n that

$$\frac{F_n - \log_2 n}{\log_2 \log n} \to -1$$

almost surely. We let F_n be the Fill level, of a p-trie with n strings and show the following lower bound – the short proof is included here for completeness. For a much larger class of random tries, F_n was studied by Pittel [22], whose results imply the bound given below.

THEOREM 2: *For $\varepsilon > 0$ and $0 < p < 1$,*

$$\lim_{n\to\infty} \mathbf{P}\left\{F_n < (1-\varepsilon)\log_{\frac{1}{\min(p,1-p)}} n\right\} = 0.$$

Proof: Without loss of generality, we assume that $p \le 1/2$. We note that

$$[F_n < l] \equiv \left[\min_{0\le i\le 2^l-1} C_{i,l} = 0\right].$$

Equivalently, by Boole's inequality, we have

$$\begin{aligned}
\mathbf{P}\{F_n < l\} \le \mathbf{P}\left\{\min_{0\le i\le 2^l-1} C_{i,l} = 0\right\} &\le \sum_{i=0}^{2^l-1} \mathbf{P}\{C_{i,l} = 0\} \\
&= 2^l \max_{0\le i\le 2^l-1} \mathbf{P}\{C_{i,l} = 0\} = 2^l\left(1 - \min_{0\le i\le 2^l-1} q_{i,l}\right)^n \\
&\le \qquad 2^l(1-p^l)^n \qquad \le 2^l e^{-np^l}.
\end{aligned}$$

This tends towards 0 with n if we take $l \sim (1-\varepsilon)\log n/\log(1/p)$ for any $\varepsilon > 0$. □

It is equally easy to show that in fact $F_n/\log_{\frac{1}{\min(p,1-p)}} n \to 1$ in probability (*see* Kruszewski [25] and Corollary 2 below).

THE HORTON-STRAHLER NUMBER

We introduce another metric, related to the Balance number. For a node u in a binary tree, we set

$$B_u^* = \begin{cases} 0 & \text{if } |u| = 0, \\ \max(B_v^* + I_{[|v|\leq|w|]}, B_w^* + I_{[|w|\leq|v|]}) & \text{if } |u| \geq 1 \text{ and } u \text{ has children } v \text{ and } w, \end{cases}$$

(see *Fig.* 5). We call B_u^* the alternate Balance number of u. It is easy to see that $B_u^* = 1$ for all leaves u. If B_n is the Balance number of any binary tree with root u, then $B_n = B_u^*$ because B_u^* is the maximum number of 1's (from the $I_{[}$'s) along any path in the tree. Note however that the Balance number of individual nodes – the $B_{i,l}$'s in the second section – are *not* equal to the quantities B_u^*.

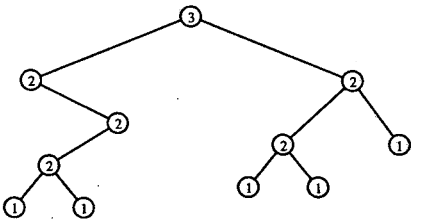

Figure 5. – Alternate Balance number labelling.

We note that the Balance number provides an upper bound on the Horton-Strahler number.

LEMMA 3: *For each binary trie with root* u, $S_u \leq B_u^*$.

Proof: For a particular tree, this follows by induction on h, the height of a node (distance from its furthest descendant leaf). At leaves u, $S_u = B_u^* = 1$. Assume that the assertion holds for all nodes of height

less than h. At height h we take a node u with children v and w. We have $S_v \leq B_v^*$, $S_w \leq B_w^*$ by assumption. If $S_v = S_w$, then, assuming $|v| \leq |w|$, we have $B_u^* \geq B_v^* + 1 \geq S_v + 1 = S_u$. If $S_v \neq S_w$, then $S_u = \max(S_v, S_w) \leq \max(B_v^*, B_w^*) \leq B_u^*$, and we are done. □

We observe that the Fill level provides a lower bound for the Horton-Strahler number.

LEMMA 4: *For each binary tree with root u,* $S_u \geq F_u$.

Proof: Trivial. □

We conclude the following tight bound on the Horton-Strahler number H_n for p-tries.

THEOREM 3: *For a p-trie with n strings,*

$$\frac{H_n}{\log_n} \to \frac{1}{\log \frac{1}{\min(p,1-p)}} \qquad \textit{in probability.}$$

Proof: The upper bound follows from Lemma 3 and Corollary 1. The lower bound follows from Lemma 4 and Theorem 2. □

This theorem together with Lemmas 3 and 4 allow us to conclude the following.

COROLLARY 2: *For a p-trie with n strings,*

$$\frac{B_n}{\log n} \to \frac{1}{\log \frac{1}{\min(p,1-p)}} \qquad \textit{in probability}$$

and

$$\frac{F_n}{\log n} \to \frac{1}{\log \frac{1}{\min(p,1-p)}} \qquad \textit{in probability.}$$

Finally, we note that as p-tries and their corresponding Patricia trees have the same Horton-Strahler numbers, our bound also hold for Patricia trees.

ACKNOWLEDGEMENTS

We would like to thank the anonymous referees for pointing out several key references.

REFERENCES

1. A. V. Aho, J. E. Hopcroft and J. D. Ullman, *Data Structures and Algorithms,* Addison-Wesley, Reading, MA, 1983.
2. D. Arquès, N. Janey and X. G. Viennot, Modélisation de la croissance et de la forme de structures arborescentes par matrice d'évolution. In *Actes de MICAD'91*, Paris, 1991, pp. 321-336.
3. L. Devroye, A note on the probabilistic analysis of Patricia trees, *Random Structures and Algorithms*, 1992, *3*, pp. 203-214
4. L. Devroye and P. Kruszewski, A note on the Horton-Strahler number for random trees, *Information Processing Letters*, 1994, *52*, pp. 155-159.
5. A. P. Ershov, On programming of arithmetic operations, *Communications of the ACM*, 1958, *1*, pp. 3-6.
6. R. Fagin, J. Nievergelt, N. Pippenger and H. R. Strong, Extendible hashing – a fast access method for dynamic files, *ACM Transactions on Database Systems*, 1979, *4*, pp. 315-344.
7. P. Flajolet, J. C. Raoult and J. Vuillemin, The number of registers required for evaluating arithmetic expressions, *Theoretical Computer Science*, 1979, *9*, pp. 99-125.
8. J. Françon, Sur le nombre de registres nécessaires à l'évaluation d'une expression arithmétique, *RAIRO Theoretical Informatics*, 1984, *18*, pp. 355-364.
9. E. H. Fredkin, Trie memory, *Communications of the ACM*, 1960, *3*, pp. 490-500.
10. W. Hoeffding, Probability inequalities for sums of bounded random variables, *Journal of the American Statistical Association*, 1963, *58*, pp. 13-30.
11. R. E. Horton, Erosional development of streams and their drainage basins; hydrophysical approach to quantitative morphology, *Bulletin of the Geological Society of America*, 1945, *56*, pp. 275-370.
12. P. Jacquet and W. Szpankowski, Analysis of digital tries with Markovian dependency, *IEEE Transactions on Information Theory*, 1991, *IT37*, pp. 1470-1475.
13. R. Kemp, The average number of registers needed to evaluate a binary tree optimally, *Acta Informatica*, 1979, *11*, pp. 363-372.
14. D. E. Knuth, *The Art of Computer Programming. Sorting and Searching*, volume 3, Addison-Wesley, Reading, MA, 1973.
15. P. Kruszewski, A probabilistic exploration of the Horton-Strahler number for random binary trees, Master's thesis, School of Computer Science, McGill University, 1993.
16. P. A. Larson, Dynamic hashing. *BIT*, 1978, *18*, pp. 184-201.
17. W. Litwin, Trie hashing. In *Proceedings of the ACM – SIGMOD Conf. MOD.,* Ann Arbor, Michigan, 1981.
18. A. Meir and J. W. Moon, Stream lengths in random channel networks, *Congressus Numerantium*, 1980, *33*, pp. 25-33.
19. A. Meir, J. W. Moon and J. R. Pounder, On the order of random channel networks, *SIAM Journal of Algebraic and Discrete Methods*, 1980, *1*, pp. 25-33.
20. J. W. Moon, On Horton's law for random channel networks, *Annals of Discrete Mathematics*, 1980, *8*, pp. 117-121.
21. J. G. Penaud, Matrice de ramification des arbres binaires, *Discrete Applied Mathematics*, 1991, *31*, pp. 1-21.
22. B. Pittel, Asymptotic growth of a class of random trees, *Annals of Probability*, 1985, *18*, pp. 414-427.
23. H. Prodinger, Solution of a problem of Yekutieli and Mandelbrot, Technical report, Technical University of Vienna, Austria, 1995.

24. A. N. STRAHLER, Hypsometric (area-altitude) analysis of erosional topology, *Bulletin of the Geological Society of America*, 1952, *63*, pp. 1117-1142.
25. J. VANNIMENUS and X. G. VIENNOT, Combinatorial Tools for the Analysis of Ramified Patterns, *Journal of Statistical Physics*, 1989, *54*, pp. 1529-1539.
26. M. VAUCHAUSSADE de CHAUMONT, *Nombre de Strahler des arbres, langages algébriques et dénombrement des structures secondaires en biologie moléculaire*, PhD thesis, Université de Bordeaux I, 1985.
27. M. VAUCHAUSSADE de CHAUMONT and X. G. VIENNOT, Enumeration of RNAs secondary structures by complexity, *Mathematics in Medicine and Biology, Lecture Notes in Biomathematics*, 1985, *57*, pp. 360-365.
28. X. G. VIENNOT, Trees everywhere. In A. Arnold ed., *Proceedings of the 15th Colloquium on Trees in Algebra and Programming, Copenhagen, Denmark, May 15-18, 1990, Lecture Notes in Computer Science*, Springer-Verlag, Berlin 1990, volume 431, pp. 18-41
29. X. G. VIENNOT, G. EYROLLES, N. JANEY and D. ARQUÈS, Combinatorial analysis of ramified patterns and computer imagery of trees. In *Proceedings of SIGGRAPH'89, Computer Graphics*, 1989, volume 23, pp. 31-40.