FURTHER RESULTS ON DIGITAL SEARCH TREES

Peter KIRSCHENHOFER and Helmut PRODINGER

Institut für Algebra und Diskrete Mathematik, Technische Universität Wien, A-1040 Vienna, Austria

Abstract. In this paper distribution results are proved on the cost of insertion in digital search trees, (binary) tries and Patricia tries. A method from the calculus of finite differences is used to achieve asymptotic results.

1. Introduction

An important class of algorithms in computer science is concerned with storing and searching for data in well-designed data structures, i.e., "digital search trees", "tries" (from information re*trie*val) and "Patricia tries" (from *p*ractical *a*lgorithm *to* *r*etrieve *i*nformation *c*oded *in* *a*lphanumeric). In the following we will present a short description of these data structures; for an extensive presentation we refer to [5, 6].

Our main purpose in this paper is the asymptotic analysis of the variances of characteristic parameters of these data structures.

Considering digital search trees, we assume that each item has a key being an infinite sequence of 0 and 1, where 0 means "go left" and 1 means "go right", until an empty space is available for the insertion of the item (cf. Fig. 1):

A: 010 . . .

B: 110 . . .

C: 111 . . .

D: 001 . . .

E: 000 . . .

Note that the order in which the keys are inserted is relevant.

(Binary) tries follow the same idea, but the items are stored in the leaves, which makes the relative order of insertion irrelevant, see Fig. 2.
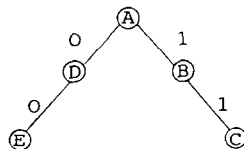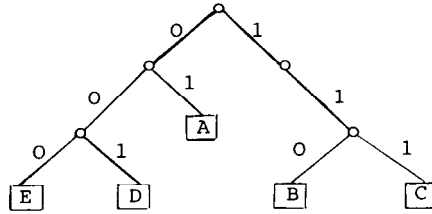

Fig. 1.

Fig. 2.

Patricia tries are constructed from tries by collapsing one-way branches on internal nodes as shown in Fig. 3.

In all three cases our parameter of interest is the cost of insertion of data.

The expectations of these parameters were determined by Knuth [2] by means of the Mellin integral transform. Flajolet and Sedgewick gave alternative derivations in [1]; they use a rather simple but very useful formula due to S.O. Rice making the whole story easier and more transparent (compare Lemma 2.1).

For the computation of the variances, different probabilistic models are meaningful which coincide for the expectations:

One possibility is to assume the uniform distribution on the set of possible keys (i.e., 0-1 sequences); work on this is in progress together with W. Szpankowski. Another possibility is to base the model on trees rather than sequences. We will choose this approach here, whence we have to go into more details.

Let $h_N^{[T]}(z)$ be the generating function with $[z^k]h_N^{[T]}(z)$ (i.e., the coefficient of z^k in $h_N^{[T]}(z)$) the expected number of external nodes at level k in the family of tries built from N records with keys from random bit streams. It should be emphasized that the expectation of the insertion cost is $(h_N^{[T]})'(1)/N$. A similar interpretation holds for Patricia tries and digital search trees (where, in the latter instance, the internal nodes have to be counted), with the obvious notations $h_N^{[P]}(z)$ respectively $h_N^{[D]}(z)$ for the corresponding generating functions.

Our model for the computation of the variance is now to consider *abstract averaging trees* the kth level of which contains $[z^k]h_N(z)$ nodes. Within this framework the variance is classically given by

$$\frac{1}{N}h_N''(1)+\frac{1}{N}h_N'(1)-\left(\frac{1}{N}h_N'(1)\right)^2.$$
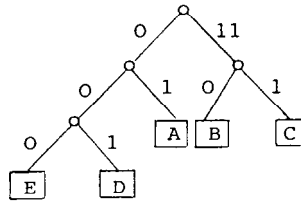

Fig. 3.

In Sections 2 to 4 of this paper we deal with the variances in all three cases; this quantity was never studied up to now. We use Rice's (or Flajolet's and Sedgewick's) method since we feel that the original approach of Knuth might be too complicated and less transparent (even though we recently learned from W. Szpankowski that a Mellin transform approach might be feasible; compare [5, 6]).

Since it is needed in the further considerations, we cite the following result.

Theorem 1.1 (Knuth; Flajolet, Sedgewick [3, 1]). *The expected value of the insertion cost for a trie respectively a Patricia trie respectively a digital search tree built from N records with keys from random bit streams is*

$$\log_2 N + \frac{\gamma}{\log 2} + \frac{1}{2} + \delta^{[T]}(\log_2 N) + O(N^{-1}),$$

respectively

$$\log_2 N + \frac{\gamma}{\log 2} + \frac{1}{2} + \delta^{[P]}(\log_2 N) + O(N^{-1}),$$

respectively

$$\log_2 N + \frac{\gamma - 1}{\log 2} + \frac{1}{2} - \alpha + \delta^{[D]}(\log_2 N) + O(N^{-1/2}).$$

Here γ is Euler's constant; $\delta^{[T]}(x) = \delta^{[P]}(x)$ and $\delta^{[D]}(x)$ are periodic functions with period 1 and very small amplitude:

$$\delta^{[T]}(x) = \frac{1}{\log 2} \sum_{k \in \mathbb{Z}, k \neq 0} \omega_k \cdot \Gamma(-\omega_k) e^{2k\pi i x},$$

$$\delta^{[D]}(x) = \frac{1}{\log 2} \sum_{k \in \mathbb{Z}, k \neq 0} \Gamma(-\omega_k) e^{2k\pi i x}$$

with $\omega_k = 1 + 2k\pi i/\log 2$; $\alpha = \sum_{k \geq 1} 1/(2^k - 1) = 1.606695 \ldots$.

So the averages are of order log N.

In Sections 2 to 4 we will prove the following result on the variances, which shows that they are of order 1.

Theorem 1.2. *The variance is asymptotic to, in the case of*
[T] *(Tries)*:

$$\frac{1}{12} + \frac{\pi^2}{6 \log^2 2} + \sigma^{[T]}(\log_2 N),$$

[P] *(Patricia tries)*:

$$\frac{1}{12} + \frac{\pi^2}{6 \log^2 2} - \frac{2}{\log 2} \cdot \sum_{n \geq 1} \frac{(-1)^{n-1}}{n(2^n - 1)} + \sigma^{[P]}(\log_2 N)$$

[D] (*Digital search trees*):

$$\frac{1}{12}+\frac{\pi^2}{6\log^2 2}+\frac{1}{\log^2 2}-\alpha-\beta+\sigma^{[D]}(\log_2 N)$$

with α *from Theorem* 1.1, $\beta=\sum_{k\geq 1}1/(2^k-1)^2$, *and the periodic functions*

$$\sigma^{[T]}(x)=\frac{2}{\log^2 2}\sum_{k\neq 0}(\Gamma(-\omega_k)-\omega_k\Gamma'(-\omega_k)-\gamma\omega_k\Gamma(-\omega_k))\,e^{2k\pi i x}-(\delta^{[T]}(x))^2,$$

$$\sigma^{[P]}(x)=\sigma^{[T]}(x)+\frac{2}{\log 2}\sum_{k\neq 0}\omega_k\Gamma(-\omega_k)(1-\xi_k)\,e^{2k\pi i x}$$

with

$$\xi_k=\sum_{n\geq 1}\binom{2k\pi i/\log 2}{n}\frac{1}{2^n-1}$$

and

$$\sigma^{[D]}(x)=\frac{2}{\log^2 2}\sum_{k\neq 0}(-\Gamma''(-\omega_k)+(1-\gamma)\Gamma(-\omega_k))\,e^{2k\pi i x}-(\delta^{[D]}(x))^2.$$

Ignoring the small fluctuations we have the following corollary.

Corollary 1.3. *The variances are roughly*

$$[T]\quad 3.5070\ldots,\qquad [P]=1.0000\ldots,\qquad [D]\quad 2.8443\ldots.$$

Concerning these numerical values we mention in passing that by means of properties of modular functions it can be proved that the constant in instance [P] differs from 1 by less than 10^{-12}; the constant in instance [D] is very close to

$$\frac{1}{24}+\frac{1}{\log^2 2}+\frac{1}{2\log 2}.$$

These results can be found in [2].

In the last section we will consider the distribution of various types of nodes in the three types of data structures. This is a continuation of the investigations of [1], where Flajolet and Sedgewick have solved an open problem of Knuth, namely to determine the expected number of internal nodes $\odot$ of the type
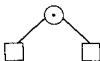
in digital search trees built from N records.

2. Tries

Let $h_N(z)$ be the generating function mentioned in the introduction (in this section always referring to tries). From [3] we know the recursion

$$h_N(z) = 2^{1-N} \sum_{k \geq 0} \binom{N}{k} z \cdot h_k(z), \quad N \geq 2,$$

$$h_0(z) = 0, \qquad h_1(z) = 1.$$

(1)

We need some more generating functions:

$$R(z) = \sum_{N \geq 0} h'_N(1) \frac{z^N}{N!}, \qquad S(z) = \sum_{N \geq 0} h''_N(1) \frac{z^N}{N!},$$

$$V(z) = e^{-z} \cdot R(z) = \sum_{N \geq 0} v_N \frac{z^N}{N!}, \qquad W(z) = e^{-z} \cdot S(z) = \sum_{N \geq 0} w_N \frac{z^N}{N!}.$$

From (1) we have

$$h''_N(1) = 2^{1-N} \sum_{k \geq 0} \binom{N}{k} (h''_k(1) + 2h'_k(1)), \quad N \geq 0,$$

and thus

$$S(z) = 2S(\tfrac{1}{2}z) e^{z/2} + 4R(\tfrac{1}{2}z) e^{z/2},$$

respectively

$$W(z) = 2W(\tfrac{1}{2}z) + 4V(\tfrac{1}{2}z).$$

From [1] we have

$$v_N = \frac{N(-1)^N}{1 - 2^{1-N}}, \quad N \geq 2,$$

so that

$$w_N = \frac{2 \cdot 2^{1-N} \cdot N \cdot (-1)^N}{(1 - 2^{1-N})^2}, \quad N \geq 2; \ w_0 = w_1 = 0.$$

By $S(z) = e^z \cdot W(z)$, we get

$$h''_N(1) = \sum_{k \geq 2} \binom{N}{k} (-1)^k \frac{2 \cdot 2^{1-k} \cdot k}{(1 - 2^{1-k})^2}.$$

(2)

The asymptotic evaluation of this alternating sum is now attacked by "Rice's method" using a classical formula for finite differences [4].

Lemma 2.1 (Nörlund [4]). *Let C be a curve surrounding the points* $2, \ldots, N$ *and* $f(z)$ *be analytic within C. Then*

$$\sum_{k \geq 2} \binom{N}{k} (-1)^k f(k) = \frac{-1}{2\pi i} \int_C [N; z] f(z) \, dz$$

with

$$[N; z] = \frac{(-1)^{N-1} N!}{z(z-1) \cdots (z-N)}.$$

Applying the lemma to expression (2) and moving the contour of integration to the left of the line with Re $z = 1$ (compare [1] for technical details), we obtain

$$h_N''(1) \sim \sum_{k \in \mathbb{Z}} \mathrm{Res}([N; z] f(z); z = \omega_k). \tag{3}$$

In order to determine the residues, we have to work out the local expansions of

$$[N; z] f(z) = [N; z] \cdot \frac{2 \cdot 2^{1-z} \cdot z}{(1 - 2^{1-z})^2} \tag{4}$$

about the poles $\omega_0 = 1$ (triple) and ω_k, $k \neq 0$ (double). For this purpose we manage to get a list of local expansions of the "ingredients": With $u = z - 1$ and $L = \log 2$ we have

$$[N; z] \sim \frac{N}{u} (1 + u(H_{N-1} - 1) + u^2 (1 - H_{N-1} + \tfrac{1}{2} H_{N-1}^2 + \tfrac{1}{2} H_{N-1}^{(2)}))$$

(H_N respectively $H_N^{(2)}$ denoting harmonic numbers)

$$2^{1-z} \sim 1 - Lu + \tfrac{1}{2} L^2 u^2,$$

$$\frac{1}{1 - 2^{1-z}} \sim \frac{1}{Lu} (1 + \tfrac{1}{2} Lu + \tfrac{1}{12} L^2 u^2), \tag{5}$$

$$\frac{1}{(1 - 2^{1-z})^2} \sim \frac{1}{L^2 u^2} (1 + Lu + \tfrac{5}{12} L^2 u^2).$$

The expansions in $u = z - \omega_k$ read

$$[N; z] \sim N^{\omega_k} [\Gamma(-\omega_k) + u(-\Gamma'(-\omega_k) + \Gamma(-\omega_k) \log N)],$$

$$2^{1-z} \sim 1 - Lu,$$

$$\frac{1}{1 - 2^{1-z}} \sim \frac{1}{Lu} (1 + \tfrac{1}{2} Lu), \tag{6}$$

$$\frac{1}{(1 - 2^{1-z})^2} \sim \frac{1}{L^2 u^2} (1 + Lu).$$

Calculating the residues of (4) we obtain, by (3),

$$h_N''(1) \sim \frac{N}{L^2}(H_{N-1}^2 + H_{N-1}^{(2)}) - \tfrac{1}{6}N$$

$$+ \frac{2N}{L^2} \sum_{k \neq 0} N^{\omega_k - 1}(\Gamma(-\omega_k) + \omega_k(-\Gamma'(-\omega_k) + \Gamma(-\omega_k) \log N)).$$

Inserting the asymptotic expansion of the harmonic numbers and using the asymptotic equivalents for $h_N'(1)$ from Theorem 1.1, we get Theorem 1.2 [T].

3. Patricia tries

We keep the notation of Section 2, but now always referring to Patricia tries. The recurrence relation for $h_N(z)$ is now

$$h_N(z) = 2^{1-N} \sum_{k \geq 0} \binom{N}{k} z h_k(z) - 2^{1-N}(z-1)h_N(z), \quad N \geq 2;$$

$$h_0(z) = h_1(z) = 1. \tag{7}$$

The functional equation for $W(z)$ reads

$$W(z) = 2W(\tfrac{1}{2}z) + 4(1 - e^{-z/2})V(\tfrac{1}{2}z) \tag{8}$$

where [1]

$$v_N = \frac{N(-1)^N}{2^{N-1} - 1}, \quad N \geq 2.$$

Hence $(N \geq 2)$,

$$w_N = \frac{2N(-1)^N}{(2^{N-1} - 1)^2} - \frac{2N(-1)^N}{2^{N-1} - 1} \sum_{n \geq 1} \binom{N-1}{n} \frac{1}{2^n - 1}.$$

Now $h_N''(1) = \sum_{k \geq 2} \binom{N}{k} w_k$ and we may apply Lemma 2.1 with

$$f(z) = \frac{2z}{(2^{z-1} - 1)^2} - \frac{2z}{2^{z-1} - 1} \sum_{n \geq 1} \binom{z-1}{n} \frac{1}{2^n - 1}.$$

We may now use the local expansions from (5) together with

$$\sum_{n \geq 1} \binom{z-1}{n} \frac{1}{2^n - 1} \sim u \cdot \sum_{n \geq 1} \frac{(-1)^{n-1}}{n(2^n - 1)} \quad \text{for } u = z - 1 \to 0$$

to obtain

$$\text{Res}([N; z]f(z); z = 1)$$

$$= N\bigg((\log_2 N)^2 + 2\frac{\gamma - L}{L} \log_2 N + \frac{\gamma^2 + \tfrac{1}{6}\pi^2}{L^2}$$

$$- \frac{2\gamma}{L} + \frac{5}{6} - \frac{2}{L} \sum_{n \geq 1} \frac{(-1)^{n-1}}{n(2^n - 1)} \bigg).$$

Next we use the local expansions from (6) and $(z \to \omega_k, k \neq 0)$

$$\sum_{n \geq 1} \binom{z-1}{n} \frac{1}{2^n - 1} \sim \sum_{n \geq 1} \binom{2k\pi i / \log 2}{n} \frac{1}{2^n - 1} =: \xi_k.$$

Thus we obtain

$$\mathrm{Res}([N; z]f(z); z = \omega_k)$$

$$= 2N\left(\frac{1}{L^2}(\Gamma(-\omega_k) + \omega_k(-\Gamma'(-\omega_k) + \Gamma(-\omega_k) \log N))\right.$$

$$\left. - \frac{1}{L}\omega_k \Gamma(-\omega_k) \cdot (1 + \xi_k)\right) e^{2k\pi i \log_2 N}.$$

The sum of the residues yields the asymptotic equivalent for $h_N''(1)$ and together with Theorem 1.1 we obtain Theorem 1.2 [P].

4. Digital search trees

From Knuth [3, p. 496] we have ($[z^k]h_N(z)$ now referring to internal nodes)

$$h_N(z) = \sum_{k \geq 0} \binom{N}{k+1}(-1)^k \prod_{0 \leq j < k}\left(1 - \frac{z}{2^j}\right) \tag{9}$$

so that, after some easy manipulations,

$$h_N''(1) = 2 \sum_{k \geq 2} \binom{N}{k}(-1)^{k-1} Q_{k-2} \cdot T(k-2) \tag{10}$$

with (compare [1])

$$Q_N = Q(1)/Q(2^{-N}) \quad \text{and} \quad Q(z) = \prod_{j \geq 1}\left(1 - \frac{z}{2^j}\right)$$

and

$$T(k) = \sum_{1 \leq j \leq k} \frac{1}{2^j - 1}.$$

The appropriate extension of $T(k)$ to $\mathbb{C}$ is

$$T(z) = \alpha - \sum_{j \geq 1} \frac{1}{2^{z+j} - 1}.$$

For the asymptotic evaluation of $h_N''(1)$ we use again Lemma 2.1 with

$$f(z) = -2 \cdot \frac{Q(1)}{Q(2^{-z+2})} \cdot T(z-2). \tag{11}$$

We start with the local expansion for $u = z - 1 \to 0$:

$$\frac{1}{Q(2^{-u+1})} = \frac{1}{1-2^{-u}} \prod_{k \geq 1} \frac{1}{1-2^{-u-k}}.$$

Now we have the following general rule for the derivatives of a product of the form

$$F(u) = \prod_{k \geq 1} \frac{1}{1-f_k(u)}:$$

$$\frac{F'(a)}{F(a)} = \sum_{k \geq 1} \frac{f'_k(a)}{1-f_k(a)},$$

$$\frac{F''(a)}{F(a)} = \left(\frac{F'(a)}{F(a)}\right)^2 + \sum_{k \geq 1} \left(\frac{f'_k(a)}{1-f_k(a)}\right)^2 + \sum_{k \geq 1} \frac{f''_k(a)}{1-f_k(a)}.$$

We apply these formulae with $f_k(u) = 2^{-u-k}$ to get $F'(0)/F(0) = -L \cdot \alpha$ and $F''(0)/F(0) = L^2(\alpha^2 + \alpha + \beta)$ with

$$\alpha = \sum_{k \geq 1} \frac{1}{2^k - 1} \quad \text{and} \quad \beta = \sum_{k \geq 1} \frac{1}{(2^k - 1)^2}.$$

From this we conclude for $u \to 0$

$$\frac{Q(1)}{Q(2^{-u})} \sim 1 - L \cdot \alpha \cdot u + \tfrac{1}{2} \cdot L^2 \cdot (\alpha^2 + \alpha + \beta) u^2.$$

We further have

$$\frac{1}{1-2^{-u}} \sim \frac{1}{Lu}(1 + \tfrac{1}{2}Lu + \tfrac{1}{12}L^2u^2),$$

so that

$$\frac{Q(1)}{Q(2^{-u+1})} = \frac{1}{1-2^{-u}} \cdot \frac{Q(1)}{Q(2^{-u})} \sim \frac{1}{Lu}(1 + L(\tfrac{1}{2} - \alpha)u + \tfrac{1}{2}L^2(\alpha^2 + \beta + \tfrac{1}{6})u^2).$$

Now we expand $T(u-1)$ for $u \to 0$:

$$T(u-1) = -\frac{1}{2^u - 1} + \alpha - \sum_{k \geq 1} \frac{1}{2^{u+k} - 1}$$

$$\sim -\frac{1}{Lu}(1 - \tfrac{1}{2}Lu + \tfrac{1}{12}L^2u^2) + \alpha - \sum_{k \geq 1} \frac{1}{2^k - 1} + L \sum_{k \geq 1} \frac{2^k}{(2^k - 1)^2} u$$

$$= -\frac{1}{Lu} + \tfrac{1}{2} + L(-\tfrac{1}{12} + \alpha + \beta)u.$$

Combining these results we get

$$\text{Res}([N; z]f(z); z = 1)$$

$$= 2N\left(\tfrac{1}{2}\alpha^2 - \tfrac{1}{2}\alpha - \tfrac{1}{2}\beta - \tfrac{1}{12} + \frac{\alpha}{L}(1 - H_{N-1})\right.$$

$$\left. + \frac{1}{L^2}(1 - H_{N-1} + \tfrac{1}{2}H_{N-1}^2 + \tfrac{1}{2}H_{N-1}^{(2)})\right).$$

The expansions for $u = z - \omega_k \to 0$ read

$$\frac{Q(1)}{Q(2^{-u-2k\pi i/L})} = \frac{Q(1)}{Q(2^{-u})} \sim 1 - L \cdot \alpha \cdot u,$$

$$\frac{1}{1 - 2^{-u-2k\pi i/L}} = \frac{1}{1 - 2^{-u}} \sim \frac{1}{L \cdot u}(1 + \tfrac{1}{2}Lu),$$

$$T(z - 2) = T\left(u - 1 + \frac{2k\pi i}{L}\right) = T(u - 1) \sim -\frac{1}{Lu} + \tfrac{1}{2}.$$
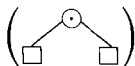
Thus we find

$$\sum_{k \neq 0} \text{Res}([N; z]f(z); z = \omega_k)$$

$$= \sum_{k \neq 0} \frac{2N^{\omega_k}}{L^2}(\log N \cdot \Gamma(-\omega_k) - \Gamma'(-\omega_k) - L\alpha\Gamma(-\omega_k))$$

$$= 2N(\log_2 N - \alpha)\delta^{[\text{D}]}(\log_2 N) - \frac{2N}{L^2}\sum_{k \neq 0}\Gamma'(-\omega_k)\,e^{2k\pi i \log_2 N}.$$

From these expansions Theorem 1.2 [D] follows immediately.

5. Distribution of various types of nodes

In solving an open problem due to Knuth, Flajolet and Sedgewick [1] have counted the average number $A_N^{[\text{D}]}$ of nodes with both sons external nodes

in digital search trees built from N records with keys from random bit streams.

Theorem 5.1 (Flajolet and Sedgewick [1])

$$A_N^{[\text{D}]} \sim N \cdot (\mu + \tau^{[\text{D}]}(\log_2 N))$$

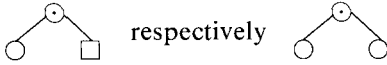*with*

$$\mu = \sum_{k \geq 1} \frac{k2^{k+1}}{1 \cdot 3 \cdots (2^k - 1)} \sum_{1 \leq j \leq k} \frac{1}{2^j - 1} + 1 - \left(\prod_{k \geq 1} \frac{2^k}{2^k - 1}\right) \cdot \left(\frac{1}{\log 2} + \alpha^2 - \alpha\right)$$

and

$$\tau^{[D]}(x) = \frac{1}{\log 2}\left(\prod_{k \geqslant 1} \frac{2^k}{2^k - 1}\right) \sum_{k \neq 0} (\omega_k - 1)\Gamma(\omega_k - 2)\, e^{2k\pi i x}.$$

The corresponding averages $B_N^{[D]}$ and $C_N^{[D]}$ of internal nodes $\odot$ of type

[tree diagram] respectively [tree diagram]

are then related to $A_N^{[D]}$ by the relations

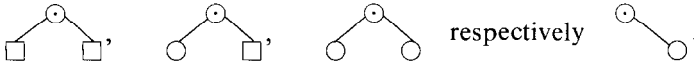$$2A_N^{[D]} + 2B_N^{[D]} = N + 1$$

(enumerating leaves) and

$$A_N^{[D]} + 2B_N^{[D]} + C_N^{[D]} = N$$

(enumerating internal nodes) so that we have the following corollary.

Corollary 5.2. $B_N^{[D]} \sim \frac{1}{2}N \cdot (1 - \mu - \tau^{[D]}(\log_2 N))$ *and* $C_N^{[D]} \sim A_N^{[D]}$.

Now we turn our attention to tries built from N records and the averages $A_N^{[T]}$, $B_N^{[T]}$, $C_N^{[T]}$ and $D_N^{[T]}$ of internal nodes $\odot$ of type

[tree diagram], [tree diagram], [tree diagram] respectively [tree diagram].

The average l_N of the total number of internal nodes is, implicitly, given in [3, p. 494]:

$$l_N \sim \frac{N}{\log 2}\left[1 + \sum_{k \neq 0} \Gamma(-\omega_k)(\omega_k - 1)\, e^{2k\pi i \log_2 N}\right]. \tag{12}$$

We have the following relations

$$2A_N^{[T]} + 2B_N^{[T]} = N \qquad \text{(enumerating leaves)},$$

$$2A_N^{[T]} + 2B_N^{[T]} + 2D_N^{[T]} = l_N + 1 \qquad \begin{array}{l}\text{(enumerating leaves of the} \\ \text{extended binary tree)},\end{array}$$

$$A_N^{[T]} + 2B_N^{[T]} + C_N^{[T]} + 2D_N^{[T]} = l_N \qquad \text{(enumerating internal nodes)}.$$

Thus we have

$$B_N^{[T]} = \tfrac{1}{2}N - A_N^{[T]}, \qquad C_N^{[T]} = A_N^{[T]} - 1,$$

$$D_N^{[T]} = \tfrac{1}{2}(l_N + 1 - N). \tag{13}$$

For $A_N^{[T]}$ we have the recurrence relation

$$A_N^{[T]} = \sum_{k \geqslant 0} \frac{1}{2^N}\binom{N}{k}(A_k^{[T]} + A_{N-k}^{[T]}), \quad N \geqslant 3,$$

$$A_0^{[T]} = A_1^{[T]} = 0, \qquad A_2^{[T]} = 1. \tag{14}$$

Using generating functions as before

$$A_N^{[T]} = \sum_{k \geq 2} \binom{N}{k} \frac{(-1)^k k(k-1)}{4(1-2^{1-k})}$$

and Rice's method can be applied to get the last theorem.

Theorem 5.3

$$A_N^{[T]} \sim \frac{N}{4 \log 2} \left(1 + \sum_{k \neq 0} \omega_k (\omega_k - 1) \Gamma(-\omega_k) \, e^{2k\pi i \log_2 N} \right).$$

The corresponding averages for Patricia tries are $A_N^{[P]} = A_N^{[T]}$, $B_N^{[P]} = B_N^{[T]}$ and $C_N^{[P]} = C_N^{[T]}$ ($D_N^{[P]} = 0$!) because of their construction from tries.

References

[1] P. Flajolet and R. Sedgewick, Digital search trees revisited, *SIAM J. Comput.* **15** (1986) 748–767.
[2] P. Kirschenhofer, H. Prodinger and J. Schoißengeier, Zur Auswertung gewisser numerischer Reihen mit Hilfe modularer Funktionen, Lecture Notes in Math. **1262** (Springer, Berlin, 1987) 108–110.
[3] D.E. Knuth, *The Art of Computer Programming, Vol. 3: Sorting and Searching* (Addison-Wesley, Reading, MA, 1973).
[4] N.E. Nörlund, *Vorlesungen über Differenzenrechnung* (Chelsea, New York, 1954).
[5] W. Szpankowski, Analysis of a recurrence equation arising in stack-type algorithms for collision-detecting channels, in: *Proc. Internat. Symp. on Computer Networking & Performance Evaluation*, Tokyo (1985) 399–412.
[6] W. Szpankowski, Some results of *v*-ary asymmetric tries, *J. Algorithms*, in press.