1987

A Characterization of Digital Search Trees from the Average Complexity Viewpoint

Wojciech Szpankowski
Purdue University, spa@cs.purdue.edu

Report Number:
87-700

A CHARACTERIZATION OF DIGITAL SEARCH TREES FROM THE AVERAGE COMPLEXITY VIEW POINT

Wojciech Szpankowski

CSD-TR-700
July 1987

A CHARACTERIZATION OF DIGITAL SEARCH TREES FROM THE AVERAGE COMPLEXITY VIEW POINT

Wojciech Szpankowski*

Department of Computer Science,
Purdue University,
West Lafayette, IN 47907, U.S.A

Abstract

This paper studies the average complexity of digital search trees from the successful search point of view. The average value of the successful search is used to evaluate the search time for a given record, the number of comparisons to insert a record, etc. The average value, however, is rather a poor measure and the need for higher moments of the successful search is obvious. For example, the variance provides information on "how well is a digital tree balanced"; the third centralized moment is a measure of the skewness property of the distribution, etc. In this paper we concentrate on an open problem: how to evaluate all moments of the successful search in an asymmetric multiway digital search tree. We prove that the m-th successful search S_n, where n is the number of stored records, satisfies $\lim_{n\to\infty} E\, S_n^m/\ln^m n = 1/h_1^m$, where h_1 is the entropy of the alphabet. In particular, it is shown that the variance of S_n is $\mathrm{var}\, S_n = c \ln n + O(1)$ for asymmetric case, and $\mathrm{var}\, S_n = O(1)$ for symmetric case (we also determine the constant). This gives a complete characterization of the digital search tree from the successful search view point.

1. INTRODUCTION

Digital searching is a well-known technique for storing and retrieving information using lexicographical (digital) structure of words. Digital search trees [2], [10], [14], [16] experience a new wave of interest due to a number of novel applications in computer science and telecommunications. For example, recent developments in the context of large external files and ideas derived from dynamic hashing (virtual hashing, dynamic hashing, extendible hashing) lead to the analysis of digital trees [5], [6], [7], [8]. Partial match retrieval of multidimensional data [8] (the grid-file, the extendible cell method) provides another application. In telecommunication, recent developments in conflict resolution algorithms [3], [9] have also brought a new interest in digital

* This research was supported in part by the National Science Foundation under grant NCR - 8702115.

trees. Some other applications are: radix exchange sort, polynomial factorization, simulation, Huffman's algorithm, etc., [2], [10], [16].

The three primary digital tree search methods are: *digital search trees, radix search tries* (shortly: tries), and *Patricia tries* [2], [10], [14], [16]. In all cases, a digital tree is built over an alphabet $A = \{\sigma_1, \ldots, \sigma_V\}$ containing V-elements. Records stored in a tree, say n of them, consists of (possibly infinite) strings (keys) from A. The *digital search* tree [2] is a data structure which leads to much improved worst case performance, by making use of the digital properties of the key. The idea is to build up a structure consisting of nodes, each node has a record containing a key and V links which point to subtrees. The branching policy on a level, say k, is based on the k-th digit (element) of a key. For example, if the k-th element of the key is σ_1, then we go to the leftmost subtree; if it is σ_2, we move to the next of the leftmost subtree, etc. However, if keys are very long, then comparison of keys at each level of the tree might be quite costly. To avoid this, in the *radix search trie* we do not store keys in tree nodes (internal nodes), but rather put all the keys in external nodes of the tree. Moreover, such a radix trie has an annoying flaw: there is "one-way branching" which leads to the creation of extra nodes in the tree. D.R. Morrison discovered a way to avoid this problem in a data structure which he named the *Patricia trie.* In such a tree, all nodes have branching degrees, greater than or equal to two. This is achieved by collapsing one-way branches on internal nodes (for more details see [2], [7], [14], [18]). Note that the number of internal nodes in the digital search tree and the Patricia trie are equal to n and $n - V + 1$, respectively. This does not hold for radix search tries. It can be proved that the average number of internal nodes is larger than n, e.g., in binary symmetric case the tree has $n/\ln 2 \sim 1.44n$ internal nodes.

Two quantities are of interest for digital trees, namely successful search and unsuccessful search. A *successful search* occurs whenever a new key added to the data structure is already in the tree. If a new key is not in the tree, then an *unsuccessful search* occurs. In this paper, we concentrate on the successful search, S_n, for a digital search tree. A complete characterization of radix search tries and Patricia tries from the successful search view point has been already obtained in [17] and [19].

To investigate the average complexity of the successful search in digital search trees , we assume that a sequence of elements from the alphabet A is an independent sequence of Bernoulli trials (*Bernoulli model*), and the probability of occurrence of an element $\sigma_i \in A$ in a key is equal to p_i, $i = 1, 2, \ldots, V$. Under these assumptions we study all moments of the successful search, S_n in a digital tree. It is shown that the m-th moment of S_n satisfies $\lim_{n\to\infty} ES_n^m/\ln^m n = 1/h_1^m$, where $h_1 = -\sum_{i=1}^{V} p_i \ln p_i$ is the entropy of the alphabet A. In particular, we prove that the variance of S_n is $(h_2 - h_1^2)h_1^{-3} \ln n + O(1)$ for an asymmetric digital tree, where $h_2 = \sum_{i=1}^{V} p_i \ln^2 p_i$. Note that this implies that the variance in the symmetric case ($p_1 = p_2 = \cdots = p_V = 1/V$) is equal to $O(1)$ (e.g. for V=2 the variance varS_n = 2.844, for V=3 we find varS_n = 1.325 and so on). To the author's knowledge, the previous analyses of the digital search tree have been restricted to binary trees and only the *average* successful search has been investigated [7], [14] (for some extension see [13]). Here we follow the approach presented in [7], but however, we differ in few aspects, as listed below.

This paper differs in some aspects from the previous analyses. We adopt a general approach to solve the problem. Namely, we first derive a general solution of a recurrence equation which governs the behavior of the successful search in digital trees. A straightforward application of this solution leads to the exact formula for $E\, S_n^m$. To obtain an asymptotic approximation, we also adopt a unified approach, that is, we first derive a general formula for some alternative sums (see Appendix) and then we use it to solve our problems. This formula generalizes Knuth's and de Bruijn's approach [14], and it is a Mellin-like technique, however, we do not explicitly use the Mellin transform.

This paper is organized as follows. In the next section, some preliminary results and summary of final results are given. We also discuss there the consequences of these results. In particular, in this section we compare moments of successful search for digital search tree, radix trie and Patricia trie. Finally, in Section 3 we prove our main results.

2. SUMMARY OF MAIN RESULTS AND DISCUSSION

Let us consider a family D_n of digital search trees with n keys (records) built over an alphabet $A = \{\sigma_1, \ldots, \sigma_V\}$. A key is a string of (possible infinite) elements from A, such that the i-th element $\sigma_i \in A$ occurs independently of other elements, and with probability p_i, $i = 1, 2, \ldots, V$, $\sum_{i=1}^{V} p_i = 1$. We study successful search, S_n, in the random family D_n of digital search trees. The m-th factorial moment of S_n is defined as follows:

$$s_n^m \stackrel{def}{=} E\{(S_n - 1)(S_n - 2), \ldots, (S_n - m + 1)\}$$

where the expectations is taken over all trees in D_n, and over all nodes in a given tree $t \in D_n$. It is shown that these moments are related to the m-th derivatives of a generating functions of D_n. Let $H_n(z)$, denote this generating function with the coefficients at z^k being the expected number of nodes (records) at level k in our family D_n .

There is no explicit formula for $H_n(z)$, but a rather sophisticated recurrence. To find it, let us denote by $j = (j_1, j_2, \ldots, j_V)$ a vector such that $j_1 + j_2 + \cdots + j_V = n$. Also let $\left(\begin{matrix} n \\ j \end{matrix}\right) \stackrel{def}{=} \left[\begin{matrix} n \\ j_1, \ldots, j_V \end{matrix}\right] = \dfrac{n!}{j!\, j_2!, \ldots, j_V!}$ be a multinomial coefficient, and let $\sum_{j_\Sigma = n} f(j_1, \ldots, j_V)$ denote a sum of $f(j_1, \ldots, j_V)$ over all j such that $j_1 + j_2 + \cdots + j_V = n$ for a given function $f(\cdot)$. Then the following recurrence on $H_n(z)$ may be established.

Lemma 1. For any n the generating function $H_n(z)$ of the random family D_n satisfies recurrence

$$H_0(z) = 0 \quad H_1(z) = 1$$
$$H_n(z) = z \sum_{\{j_\Sigma = n-1\}} \binom{n-1}{j} p_1^{j_1}, \ldots, p_V^{j_V} [H_{j_1}(z) + \cdots + H_{j_V}(z)] + 1 \tag{2.1}$$

Proof. Consider V subtrees of the root, each with $j_1, j_2, \ldots, j_V$ keys, $j_1 + j_2 + \cdots + j_V = n-1$. Then, for a given tree $t \in D_n$

$$H_n(z) = z\ [H_{j_1}(z) + \cdots + H_{j_V}(z)]$$

The first term z in the above shows the fact that the subtrees are one level below the root. Taking now the expectation of the last recurrence over all trees in D_n, and noting that in our Bernoulli model the probability of $j_1, \ldots, j_V$ keys in the subtrees is equal to $\binom{n-1}{j} p_1^{j_1}, \ldots, p_V^{j_V}$, we finally obtain (2.1).

□

Now we establish relationship between s_n^m and the m-th derivative of $H_n(z)$ at z=1. Let L_n denote an internal path in a digital tree $t \in D_n$, that is, the sum of all paths from the root to all nodes. We generalize the definition of L_n as follows. Let $S_n(i)$ be a path from the root to the i-th node. For a given integer m we define

$$L_n^m = \sum_{i=1}^{n} S_n(i)[S_n(i) - 1][S_n(i) - 2], \ldots, [S_n(i) - m + 1]$$

and let $l_n^m = EL_n^m$. The quantity l_n^m is not exactly the m-th factorial moment of L_n, but it is closely related to it. We call l_n^m the m-th semi-factorial moment of the internal path length.

Denote now by $H_n^{(m)}(1)$ the m-th derivative of $H_n(z)$ at z=1. Then the following is easy to establish (see [17])

Property 1. For integers n and m the below relationships hold

$$H_n(1) = n \qquad l_n^m = H_n^{(m)}(1) \tag{2.2}$$
$$s_n^m = l_n^m / n \tag{2.3}$$

□

Using Lemma 1 and (2.2), we derive a recurrence equation for l_n^m, hence by (2.3) also on s_n^m. We shall work at the beginning with l_n^m. For simplicity of the presentation, assume that $V = 2$ and $p_1 = p, p_2 = 1 - p_1 = q$, however, all results can be trivially extended to V-ary asymmetric digital search trees. From (2.1) and (2.2), for $m=1$ we find immediately that

$$l_n^1 = n-1 + \sum_{k=0}^{n-1} \binom{n-1}{k} p^k q^{n-1-k} [l_k^1 + l_{n-1-k}^1] \qquad (2.4)$$

Computing the second derivative of $H_n(z)$ one shows that

$$l_n^2 = 2[l_n^1 - (n-1)] + \sum_{k=0}^{n-1} \binom{n-1}{k} p^k q^{n-1-k} [l_k^2 + l_{n-1-k}^2]. \qquad (2.5)$$

Note that (2.4) and (2.5) is a system of recurrences, i.e., to find l_n^2 we need l_n^1. Generalizing the above, we can prove that

Lemma 2. For any integers m and n, the m-th semi-factorial moment of L_n satisfies the following recurrence

$$l_0^m = l_1^m = 0$$

$$l_n^m = m! \sum_{k=1}^{m} (-1)^{m-k} \frac{l_n^{k-1}}{(k-1)!} + \sum_{\{j_1 = n-1\}} \binom{n-1}{j} p_1^{j_1}, \ldots, p_V^{j_V} [l_{j_1}^m + \cdots + l_{j_V}^m] \qquad (2.6)$$

where in (2.6) we have defined $l_n^0 = n-1$.

Proof. The proof uses induction arguments applied to (2.3), and is left to the reader.

□

As noted before, (2.6) is a system of recurrences. To compute l_n^m we need l_n^1, $l_n^2, \ldots, l_n^{m-1}$ from the previous recurrences. Note also that (2.6) has a common pattern and the recurrences differ only by the first term in (2.6) which we call the *additive term* and denote by a_n. To present a unified analysis of all moments of S_n we first solve a general recurrence of type (2.6).

Solution of a recurrence equation

Let $x_0, x_1, \ldots, x_n$ be a sequence of numbers satisfying the following recurrence equation

$$\text{given } x_0, x_1$$

$$\text{solve} \quad x_n = a_n + \sum_{j_\Sigma = n-1} \binom{n-1}{\mathbf{j}} p_1^{j_1} \cdots p_V^{j_V} [x_{j_1} + \cdots + x_{j_V}], \quad n \geq 2 \tag{2.7}$$

where a_n is any sequence of numbers. We call a_n an additive term of the recurrence (2.7). To solve (2.7) we introduce the so called *binomial inverse relations*. Let us, for a given sequence a_n, define a new sequence $\hat{a}_n$ as

$$\hat{a}_n = \sum_{k=0}^{n} (-1)^k \binom{n}{k} a_k \qquad a_n = \sum_{k=0}^{n} (-1)^k \binom{n}{k} \hat{a}_k \tag{2.8}$$

(The second equation justifies the name binomial inverse relations). For more details, see Riordan [23]. We prove

Theorem 1. The recurrence (2.7) possesses the following solution

$$x_n = x_0 + n x_1 - \sum_{k=2}^{n} (-1)^k \binom{n}{k} R_{k-2} \tag{2.9a}$$

where

$$R_n = Q_n \sum_{i=1}^{n+1} [\hat{a}_i - \hat{a}_{i+1} - A_0] Q_{i-1}^{-1} \tag{2.9b}$$

$$Q_n = \prod_{j=2}^{n+1} (1 - \sum_{k=1}^{V} p_k^j), \quad Q_0 = 1 \tag{2.9c}$$

and $A_0 = a_1 + 2x_0 - x_1$.

Proof. Multiply both sides of (2.7) by $z^{n-1}/(n-1)!$ and sum from 0 to infinity. Let $X(z)$ and $A(z)$ represent the exponential generating functions of x_n and a_n respectively. Then, (2.7) reduces to the below equation

$$X'(z) = A'(z) - A_0 + \sum_{i=1}^{V} X(p_i z) e^{(1-p_i)z} \tag{2.10}$$

Introducing $Y(z) = X(z) e^{-z}$ one transforms (2.10) into

$$Y'(z) + Y(z) = B(z) + \sum_{i=1}^{V} Y(p_i z) \tag{2.11}$$

where $B(z) = [A'(z) - A_0]e^{-z}$. Equating coefficients in (2.11) we find a recurrence

$$y_{n+1} + y_n = b_n + y_n \sum_{i=1}^{V} p_i^n \qquad (2.12)$$

where $b_n = (-1)^n(\hat{a}_n - \hat{a}_{n+1} - A_0)$. The solution of (2.12) is

$$y_n = (-1)^{n-1} \sum_{i=1}^{n-1} (-1)^i \; b_i \prod_{j=i+1}^{n-1} (1 - \sum_{k=1}^{V} p_k^j)$$

But $x_n = \sum_{k=0}^{n} \binom{n}{k} y_k$, hence after some algebra (2.9) follows.

□

Note also that from the definition of the inverse relations (2.8) and solution (2.9) one easily proves

Corollary 1. If $x_0 = x_1 = 0$, then the inverse relation, $\hat{x}_n$, of the solution x_n is given by

$$\hat{x}_n = R_{n-2} \qquad n \geq 2 \qquad (2.13)$$

where R_n is defined in (2.9b).

□

To find an asymptotic approximation for (2.9c) we need to evaluate the alternative binomial sum in (2.9c). We present below a general approach to find such an approximation. Let

$$F_m(n) = \sum_{k=m}^{n} (-1)^k \binom{n}{k} f(k) \qquad (2.14)$$

where $f(k)$ is a function of k such that an analytical continuation of $f(k)$ to a complex function $f(z)$ exists. Then in the Appendix * we prove

Theorem 2. If $f(z)$ is analytical left to the line $(\frac{1}{2} - m - i\infty,\ \frac{1}{2} - m + i\infty)$, then

$$F_m(n) = \frac{1}{2\pi i} \int_{\frac{1}{2} - m - i\infty}^{\frac{1}{2} - m + i\infty} \Gamma(z) f(-z) n^{-z} dz + e_n \qquad (2.15)$$

where

The detailed proof of Theorem 2, with necessary conditions on $f(z)$, is given in [20]. In the final version of this paper the proof can be deleted, if required.

$$e_n = O(n^{-1}) \frac{1}{2\pi i} \int_{\frac{1}{2}-m-i\infty}^{\frac{1}{2}-m+i\infty} z\,\Gamma(z) f(-z) n^{-z}\, dz$$

and $\Gamma(z)$ is the gamma function [11], [22].

□

Note that $e_n = o(n)$. In our case we prove that $e_n = O(1)$. Evaluation of the sum (2.14) by formula (2.15) is routine. We appeal to Cauchy's theorem. The integral in (2.15) is equal to minus the sum of residues *right* to the line of integration. For details see Section 3 and [14], [17], [20].

Now we are ready to present out the main results. Using Lemma 2, Property 1 and Theorem 1, we prove in the next section that the exact solution for the m-th factorial moment of S_n is given by

Proposition 1. For any m and n, the m-th factorial moment, s_n^m, of the successful search S_n can be expressed as

$$s_n^m = \frac{(-1)^{m-1}\, m!}{n} \sum_{k=2}^{n} (-1)^k \begin{bmatrix} n \\ k \end{bmatrix} Q_{k-2}\, T_{k-2}^{(m-1)} \qquad (2.16)$$

where Q_n is defined in (2.9c) and $T_m^{(m)}$ is computed by the following recursion : $T_n^{(0)} = 1$ and for $m > 0$

$$T_n^{(m)} = \sum_{i=2}^{n+1} T_{i-2}^{(m-1)} \frac{\sum_{k=1}^{V} p_k^i}{1 - \sum_{k=1}^{V} p_k^i} \qquad (2.17)$$

□

Using (2.16) and Theorem 2, we obtain asymptotics for S_n^m.

Proposition 2. The average successful search, ES_n, for large n is given by

$$ES_n = \frac{1}{h_1}\{\ln n + \gamma - 1 + \frac{h_2}{2h_1} - \theta_1 + f_1(n)\} + \frac{1}{h_1}\,\frac{\ln n}{n} + O(n^{-1}) \qquad (2.18)$$

where $h_n = (-1)^n \sum_{k=1}^{V} p_k \ln^n p_k$, $\gamma = 0.577$ is the Euler constant, $f_1(n)$ is a fluctuating function with very small amplitude, and

$$\theta_1 = -\sum_{k=1}^{\infty} \frac{\sum_{i=1}^{V} p_i^{k+1} \ln p_i}{1 - \sum_{i=1}^{V} p_i^{k+1}} \tag{2.19}$$

(ii) The variance of S_n for large n becomes

$$\mathrm{var}\, S_n = \frac{h_2 - h_1^2}{h_1^3} \ln n + C + F(n) + O(n^{-1} \ln^2 n) \tag{2.20}$$

where the constant C is computed in Section 3 (see (3.25) - (3.28)). In particular, for symmetric case $h_2 = h_1 = \ln^2 V$ and the first term in (2.20) disappears. In this case $\mathrm{var}\, S_n = C$, that is

$$\mathrm{var} S_n = \frac{1}{12} + \frac{1}{\ln^2 V}\left[\frac{\pi^2}{6} + 1\right] - \alpha - \beta + F(n) + O(n^{-1} \ln^2 n) \tag{2.21}$$

where

$$\alpha = \sum_{j=1}^{\infty} \frac{1}{V^j - 1} \qquad \beta = \sum_{j=1}^{\infty} \frac{1}{(V^j - 1)^2} \tag{2.22}$$

and $F(n)$ is a fluctuating function with a very small amplitude.

(iii) The m-th moment of S_n, ES_n^m, satisfies

$$\lim_{n \to \infty} \frac{ES_n^m}{\ln^m n} = \frac{1}{h_1^m} \tag{2.23}$$

□

Propositions 1 and 2 complete the classification of asymmetric digital trees, that is, radix tries [17], Patricia tries [19] and digital search trees (this paper). For all three digital trees, the mean value of S_n is $\frac{1}{h_1} \ln n + O(1)$ with different constants $O(1)$. The variance, $\mathrm{var} S_n$, for an asymmetric case is $O(\ln n)$, however, in the symmetric case $\mathrm{var} S_n = O(1)$ and the constant differs significantly. Let S_n^D, S_n^T, S_n^P, denote the successful searches for digital search trees, radix tries and Patricia tries respectively. Then by (2.18) and results from [17] [19], one shows

$$ES_n^D - ES_n^T = -(1 + \theta_1)/h_1 < 0 \qquad (2.24a)$$

$$ES_n^D - ES_n^P = -(1 + \theta_1 - \bar{h})/h_1 < 0 \qquad (2.24b)$$

where $\bar{h} = -\sum_{i=1}^{V} p_i \ln(1 - p_i)$. Therefore, the best constant in ES_n is achieved for the digital search tree. On the other hand, the best variance of S_n is for Patricia tries. To see it, note that for symmetric case, the variance for radix tries and Patricia tries is given by [13] [17] [18] [19].

$$\begin{aligned} \operatorname{var} S_n^T &= \frac{1}{12} + \frac{\pi^2}{6 \ln^2 V} \\ \operatorname{var} S_n^P &= \frac{1}{12} + \frac{\pi^2}{6 \ln^2 V} - \frac{2}{\ln V} \left\{ \prod_{l=1}^{\infty} (1 + \frac{1}{V^l} \right\} \end{aligned} \qquad (2.25)$$

The table below compares the variances in the symmetric case for the above three trees.

V	$\operatorname{var} S_n^T$	$\operatorname{var} S_n^P$	$\operatorname{var} S_n^D$
2	3.507	1.000	2.844
3	1.446	0.630	1.325
4	0.939	0.500	0.923
5	0.718	0.430	0.738

Note that, as we have argued in [17], the variance of S_n indicates *how well a tree is balanced.* Thus, the above shows that the Patricia trie is the best balanced digital tree, and the regular trie is the worst one. Note, however, that Patricia tries required $2n - V + 1$ nodes compared to n nodes for digital search trees. But, the comparison of keys at each level of the digital search tree might be quite costly if keys are very long. On the other hand, the Patricia trie is the most sophisticated digital tree, since additional pointers are required to indicate over how many digits the search procedure must skip to locate the next "branching" digits in an inserted key.

3. ANALYSIS

In this section we prove Propositions 1 and 2. Hereafter, for simplicity, we consider only asymmetric binary digital trees, that is, $V=2$, $p=p_1$ and $p_2=1-p_1=q$. The extension to the V-ary digital trees is trivial.

3.1. The average value of the successful search

The average value of the internal path length is given by recurrence (2.4) which falls into (2.7) with $a_n = n-1, x_0 = x_1 = 0$. Note that $\hat{a}_n = -\delta_{n,1} - \delta_{n,0}$, where $\delta_{n,k}$ is the Kronecker delta [14], [21]. Hence, by Theorem 1, we find

$$l_n^1 = \sum_{k=2}^{n} (-1)^k \begin{bmatrix} n \\ k \end{bmatrix} Q_{k-2} \tag{3.1}$$

and by Corollary 1

$$\hat{l}_n^1 = Q_{n-2} \tag{3.2}$$

where Q_n is given by (2.9c). This proves Proposition 1 for $m=1$.

To find an asymptotic approximation for l_n^1 we apply Theorem 2. Therefore, we need to define a complex function $Q(z)$ that extends Q_k. Flajolet and Sedgewick [7] dealt with Q_k in the analysis of binary symmetric digital tree. (Note that they used Rice's method to evaluate l_n^1, while we adopt approximations from Theorem 2). Unfortunately, the extension $Q(z)$ proposed for the symmetric case cannot here be directly applied. Thus, we propose to use

$$Q(z) = \frac{P(0)}{P(z)} \tag{3.3a}$$

where

$$P(z) = \prod_{j=2}^{\infty} (1 - p^{z+j} - q^{z+j}) \tag{3.3b}$$

It is easy to see that $Q(z)$ for z nonnegative integers coincides with Q_k, and $Q(0) = Q_0 = 1$. Note, however, that (3.3) is a proper analytical extension of Q_k if and only if the product in (3.3b) is convergent. But this holds if the following series is convergent [11]

$$\sum_{j=1}^{\infty} (p^{z+1} + q^{z+1}) = p^z \frac{p^2}{1-p} + q^z \frac{q^2}{1-q}$$

Hence, (3.3) is an appropriate extension.

Now using (3.2), (3.3) and Theorem 2, we find

$$l_n^1 = \int_{(-3/2)} \Gamma(z) n^{-z} Q(-z-2) dz + O(1) \tag{3.4}$$

where $\int_{(c)} f(\cdot)$ stands for $\frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} f(\cdot)$. Evaluation of (3.4) is standard, and appeal to the residue theorem. Note that the function under the integral, say $g(z)$, is analytical right to the line $(-3/2 - i\infty, -3/2 + i\infty)$ except $z_0 = 0$, $z_{-1} = -1$ (singularities of the gamma functions) and zeros $z_k(j)$, of $P(-z-2)$, that is

$$z_k(j) = j - 2 + iy_k \quad k = 0 \pm 1, \ldots, \; j = 1, 2, \ldots, \tag{3.5}$$

where $i\,y_k$ is the imaginary part of $z_k(j)$ with $y_0=0$. But $z_0 = z_0(2)$ and $z_{-1} = z_0(1)$, hence z_0 and z_{-1} are double poles. We shall see that the main contribution comes from z_{-1} (and also z_0), while $z_k(j)$, $k \neq 0$ give a fluctuating function with very small amplitude (see [6]–[8], [13–[14], [17]–[19]). We denote this function by $f_1(n)$.

To compute residue at $z_{-1} = -1$, res $g(z_{-1})$, we need the Taylor expansion of the functions under the integral at $z_{-1} = -1$. But [1], [11], [22] for $w = z + 1$,

$$\begin{aligned} \Gamma(z) &= -w^{-1} + (\gamma - 1) + O(w) \\ n^{-z} &= n(1 - w \ln n) \end{aligned} \tag{3.6}$$

To find expansion of $Q(z)$, we apply Lemma 2 from [7] and the fact [17]

$$\frac{1}{1 - p^{-z} - q^{-z}} = -\frac{w^{-1}}{h_1} + \frac{h_2}{2h_1^2} + O(w)$$

where $h_n = (-1)^n \sum_{i=1}^{V} p_i (\ln p_i)^n$. Then

$$\begin{aligned} Q(-z-2) = \frac{P(0)}{P(-z-2)} &= \frac{1}{1 - p^{-z} - q^{-z}} \frac{P(0)}{P(-1-z)} = \\ -\frac{w^{-1}}{h_1} - \frac{\theta_1}{h_1} &+ \frac{h_2}{2h_1^2} + w \frac{\theta_1 h_2}{2h_1^2} + O(w^2) \end{aligned} \tag{3.7}$$

where θ_1 is defined in (2.19). Multiplying the above, and evaluating the coefficient at w^{-1} we obtain

$$\text{res } g(z_{-1}) = -\frac{n}{h_1}\{\ln n + h_2/2h_1 + \gamma - 1 - \theta_1\}$$

In a similar way, one proves that

$$\text{res } g(z_0) = -\frac{1}{h_1}\{\ln n - \gamma + h_2/2h_1 - \ln p - \ln q + \theta_1\}$$

This, with the additional contribution coming from $z_k(j)$, $k \neq 0$ (that is, the function $f_1(n)$), proves Proposition 2 (i).

3.2 The variance of the successful search

The computation of l_n^2, and therefore $\text{var} S_n = [l_n^2 + l_n^1 - (l_n^1)^2]/n$, is much more intricate. Note first that (2.4) and (2.5) imply that

$$l_n^2 = 2(X_n - l_n^1) \tag{3.8}$$

where X_n satisfies the following recurrence

$$X_n = l_n^1 + \sum_{k=0}^{n-1} \binom{n-1}{k} p^k q^{n-1-k} [X_n + X_{n-1-k}] \quad n \geq 2 \tag{3.9}$$

The recurrence (3.9) falls into (2.7) with $a_n = l_n^1$. By (3.2) we know that $l_n^1 = Q_{n-2}$, therefore, after some algebra we find

$$X_n = l_n^1 - \sum_{k=2}^{n} \binom{n}{k} (-1)^k Q_{k-2} T_{k-2}^{(1)} \tag{3.10}$$

where

$$T_n^{(1)} = \sum_{i=2}^{n+1} \frac{p^i + q^i}{1 - p^i - q^i} \tag{3.11}$$

and, finally, (3.8) and (3.10) imply

$$l_n^2 = -2 \sum_{k=2}^{n} \binom{n}{k} (-1)^k Q_{k-2} T_{k-2}^{(1)} \tag{3.12}$$

which proves Proposition 1 for $m = 2$.

To estimate asymptotics for l_n^2, we apply Theorem 2. Hence, an analytical continuation of $T_k^{(1)}$ to a complex function $T^{(1)}(z)$ is necessary. To accomplish that we adopt the "mechanical derivation" suggested in [7]. Note that by (3.11)

$$T_{n+1}^{(1)} = T_n^{(1)} + \frac{p^{n+2} + q^{n+2}}{1 - p^{n+2} - q^{n+2}}$$

Replacing n by z in the above, we obtain

$$T^{(1)}(z) = T^{(1)}(z+1) - \frac{p^{z+2} + q^{z+2}}{1 - p^{z+2} - q^{z+2}} \tag{3.13}$$

This is a recurrence equation for which a solution becomes

$$T^{(1)}(z) = T^{(1)}(\infty) - \sum_{i=2}^{\infty} \frac{p^{z+i} + q^{z+i}}{1 - p^{z+i} - q^{z+i}}$$

Define $T^{(1)}(\infty) = \alpha$. Then, by (3.11)

$$\alpha = \sum_{i=2}^{\infty} \frac{p^i + q^i}{1 - p^i - q^i} \tag{3.14}$$

and finally

$$T^{(1)}(z) = \alpha - \sum_{i=2}^{\infty} \frac{p^{z+i} + q^{z+i}}{1 - p^{z+i} - q^{z+i}} \tag{3.15}$$

To prove that (3.15) is really the analytical continuation of $T_n^{(1)}$ one must show that the series in (3.15) is convergent. This is easy using the same arguments as applied in Sec. 3.1.

Now we are ready to compute asymptotics for l_n^2. By Theorem 2

$$\sum_{k=2}^{n} \left[\begin{matrix} n \\ k \end{matrix} \right] (-1)^k Q_{k-2} T_{k-2}^{(1)} = \int_{(-3/2)} \Gamma(z) n^{-z} \frac{P(0)}{P(-z-2)} T^{(1)}(-z-2) dz + O(1) \tag{3.16}$$

We evaluate the integral by the residue theorem. Note that, as before, we have singularities at $z_0 = 0$, $z_{-1} = -1$ and $z_k(j)$, however now z_0 and z_{-1} are triple poles. The main contribution comes from $z_{-1} = -1$. Therefore we need the Taylor expansion of the functions under the integral *up to three terms*. The previous formula for $\Gamma(z)$, n^{-z} may be applied (with one additional term). For $T^{(1)}(z)$, we split the function into two terms, that is, $T^{(1)}(-z-2) = T_1(z) + T_2(z)$ where

$$T_1(z) = 1 - \frac{1}{1 - p^{-z} - q^{-z}} \tag{3.17a}$$

$$T_2(z) = \alpha - \sum_{j=3}^{\infty} \frac{p^{-z-2+i} + q^{-z-2+i}}{1 - p^{-z-2+i} - q^{-z-2+i}} \tag{3.17b}$$

Naturally, for $w = z+1$

$$T_2(z) = -\theta_2 w + O(w^2)$$

where

$$\theta_2 = -\sum_{k=1}^{\infty} \frac{p^k \ln p + q^k \ln q}{(1 - p^{k+1} - q^{k+1})^2} \tag{3.18}$$

On the other hand, the Taylor expansion of $T_1(z)$ is given by (see [17]).

$$T_1(z) = \frac{w^{-1}}{h_1} + 1 - \frac{h_2}{2h_1^2} + w\left[\frac{h_2^2}{4h_1^3} - \frac{h_3}{6h_1^2}\right] + O(w^2)$$

Therefore,

$$T^{(1)}(-z-2) = \frac{w^{-1}}{h_1} + \left[1 - \frac{h_2}{2h_1^2}\right] + w\left[\frac{h_2^2}{4h_1^3} - \frac{h_3}{6h_1^2} - \theta_2\right] + O(w^2) \tag{3.19}$$

We also need three terms in the Taylor expansion of $Q(-z-2)$. Note that

$$Q(-z-2) = \frac{P(0)}{P(-z-2)} = \frac{P(0)}{1 - p^{-z} - q^{-z}} \prod_{j=1}^{\infty} (1 - p^{-z+j} + q^{-z+j})^{-1} \tag{3.20}$$

We apply to (3.20) the following lemma, which extends Lemma 2 from [7].

Lemma 3. Let $F(z) = \prod_{j \in S} [1 - f_j(z)]^{-1}$, where S is an index set, and let $F(a)$ exist for some real a. Then

$$F(z) = F(a)\left\{ 1 + (z-a) \sum_{j \in S} \frac{f_j'(a)}{1 - f_j(a)} + \frac{(z-a)^2}{2}\left[\left[\sum_{j \in S} \frac{f_j'(a)}{1 - f_j(a)}\right]^2 + \sum_{j \in S} \frac{f_j''(a)(1 - f_j(a)) + [f_j'(a)]^2}{[1 - f_j(a)]^2}\right]\right\} + O((z-a)^3) \tag{3.21}$$

Proof. Use logarithm derivative for $G(z) = \prod_{j \in S} g_j(z)$.

□

A straightforward application of Lemma 3 to (3.20) gives

$$Q(-z-2) = -\frac{w^{-1}}{h_1} + \left[\frac{h_2}{2h_1^2} - \frac{\theta_1}{h_1} \right] + w \left[\frac{h_3}{6h_1^2} - \frac{h_2^2}{4h_1^3} - \frac{\theta_1^2 + \beta_1 + \beta_2}{h_1} + \frac{\theta_1 h_2}{2h_1^2} \right] + O(w^2) \quad (3.22)$$

where

$$\beta_1 = \sum_{k=1}^{\infty} \frac{p^{k+1}\ln^2 p + q^{k+1}\ln^2 q}{1 - p^{k+1} - q^{k+1}} \quad (3.23a)$$

$$\beta_2 = \sum_{k=1}^{\infty} \left\{ \frac{p^{k+1}\ln p + q^{k+1}\ln q}{1 - p^{k+1} - q^{k+1}} \right\}^2 \quad (3.23b)$$

Finally, multiplying (3.19), (3.22) and appropriate formula on $\Gamma(z)$ and n^{-z}, and identifying the coefficient at w^{-1}, we prove

$$l_n^2 = \frac{n \ln^2 n}{h_1^2} + \frac{2n \ln n}{h_1} \left[\frac{h_2}{h_1} - 1 - \frac{\theta_1}{h_1} + \frac{\gamma - 1}{h_1} \right] + 2nA + f_2(n) + O(\ln^2 n) \quad (3.24)$$

where $f_2(n)$ is a fluctuating function coming from the contribution of $z_k(j)$, $k \neq 0$, and $O(\ln^2 n)$ is the contribution from $z_0 = 0$. The constant A is equal to

$$A = \frac{(\gamma - 1)(\varepsilon_1 - \delta_1)}{h_1} - \frac{\gamma_1}{h_1^2} - \varepsilon_1\delta_1 + \frac{\delta_2}{h_1} - \frac{\varepsilon_2}{h_1} \quad (3.25)$$

where

$$\varepsilon_1 = \frac{h_2}{2h_1^2} - \frac{\theta_1}{h_1} \quad ; \quad \varepsilon_2 = \frac{h_3}{6h_1^2} - \frac{h_2^2}{4h_1^2} - \frac{\theta_1^2 + \beta_1 + \beta_2}{2h_1} + \frac{\theta_1 h_2}{2h_1^2} \quad (3.26a)$$

$$\delta_1 = 1 - \frac{h_2}{2h_1^2} \quad ; \quad \delta_2 = -\frac{h_3}{6h_1^2} + \frac{h_2^2}{4h_1^2} - \theta_2 \quad (3.26b)$$

and γ_1 is the coefficient at w in the Taylor expansion of the gamma function. Using results from [4], we can prove that

$$\gamma_1 = -\frac{1}{2} \left[\frac{\pi^2}{6} + 1 - (1 - \gamma)^2 \right] \quad (3.26c)$$

The rest is easy. Note that $\mathrm{var} S_n = \frac{1}{n}\{l_n^2 + l_n^1 - (l_n^1)^2\}$. In Section 3.1, we have proved that $l_n^1 = \frac{\ln n}{h_1} + B$, where

$$B = \frac{1}{h_1}(\gamma - 1 + \frac{h_2}{2h_1} - \theta_1 + f_2(n)) \tag{3.27}$$

After some algebra, one proves Proposition 2(ii) formula (2.20) where the constant C is equal to

$$C = 2A + B - B^2 \tag{3.28}$$

with A and B given by (3.25) and (3.27). In the symmetric case, the constant C simplifies to (2.21) by taking into account the fact that $h_k = \ln^k V$ and $\beta_2 = h_1^2 \beta$ where β is given by (2.2).

3.3 The higher moments

The proof of Proposition 1 for general m is by induction. To simplify the presentation, we show in this subsection how the proof goes for $m = 3$, and left the details of the induction for the interested reader.

By Lemma 2 with $m = 3$, we find

$$l_n^3 = 6[l_n^2/2 - l_n^1 + (n-1)] + \sum_{k=0}^{n-1} \binom{n-1}{k} p^k q^{n-1-k} [l_k^3 + l_{n-1-k}^3]$$

Let X_n be defined as: $X_0 = X_1 = 0$ and for $n \geq 2$

$$X_n = l_n^2 + \sum_{k=0}^{n-1} \binom{n-1}{k} p^k q^{n-1-k} [X_k + X_{n-1-k}] \tag{3.29a}$$

Then, by the above and (2.5)

$$l_n^3 = 6[X_n/2 - l_n^2/2] \tag{3.29}$$

To solve (3.29a) we apply Theorem 1. We need to compute R_n given by (2.9b). In our case, $a_n = l_n^2$ and $\hat{a}_n = \hat{l}_n^2$. By (3.12) $\hat{l}_n^2 = -2\, Q_{k-2} T_{k-2}^{(1)}$. Then

$$R_n = 2T_n^{(1)} Q_n - 2Q_n \sum_{i=2}^{n+1} T_{i-2}^{(1)} \frac{p^i + q^i}{1 - p^i - q^i} \tag{3.30}$$

Finally, by (3.29) and Theorem 1

$$l_n^3 = 6 \sum_{k=2}^{n} (-1)^k \begin{bmatrix} n \\ k \end{bmatrix} Q_{k-2}\, T_{k-2}^{(2)} \tag{3.31a}$$

where

$$T_{k-2}^{(2)} = \sum_{i=2}^{n+1} T_{i-2}^{(1)} \frac{p^i + q^i}{1 - p^i - q^i} \tag{3.31b}$$

This proves Proposition 1 with $m = 3$. Extension to general m is simple and left to the reader.

Asymptotic analysis of (3.31a), or in general (2.16) for any m, requires analytical continuation of $T_n^{(m)}$ defined as (see (2.17)).

$$T_n^{(m)} = \sum_{i=2}^{n+1} T_{i-2}^{(m-1)} \frac{p^i + q^i}{1 - p^i - q^i} \tag{3.32}$$

Let $T^{(m)}(z)$ be such an analytical continuation. Then, arguing as in (3.13), we find the following recurrence

$$T^{(m)}(z) = T^{(m)}(z+1) - T^{(m-1)}(z) \frac{p^{z+2} + q^{z+2}}{1 - q^{z+2} - q^{z+2}}$$

This has a solution

$$T^{(m)}(z) = T^{(m)}(\infty) - \sum_{i=2}^{\infty} T^{(m-1)}(z-2+i) \frac{p^{z+i} + q^{z+i}}{1 - p^{z+i} - q^{z+i}} \tag{3.33}$$

But, by (3.32) $T^{(m)}(\infty) = \alpha^{(m)}$ where

$$\alpha^{(m)} = \sum_{i=2}^{\infty} T_{i-2}^{(m-1)} \frac{p^i + q^i}{1 - p^i - q^i} \tag{3.34a}$$

hence

$$T^{(m)}(z) = \alpha^{(m)} - \sum_{i=2}^{\infty} T^{(m-1)}(z-2+i) \frac{p^{z+i} + q^{z+i}}{1 - p^{z+i} - q^{z+i}} \tag{3.34b}$$

Note that $T^{(0)}(z) = 1$, $T^{(1)}(z)$ is given by (3.15) and $\alpha^{(1)} = \alpha$ as in (3.14). Thus, analytical continuation of $T_n^{(m)}$ is done.

To prove Proposition 2(iii) formula (2.23), we apply Theorem 2, hence

$$l_n^m = (-1)^{m-1}\, m! \int_{(-3/2)} \Gamma(z) n^{-z} \frac{P(0)}{P(-z-2)} T^{(m-1)}(-z-2) dz + O(1) \tag{3.35}$$

We consider only residues at $z_{-1} = -1$. From the previous analysis, we know that $\Gamma(z) = -w^{-1} + O(1)$ and $P(0)/P(-z-2) = -w^{-1}/h_1 + O(1)$. Naturally, $n^{-z} = n \sum_{k=0}^{m} (-1)^k \frac{w^k \ln^k n}{k!} + O(w^{m+1})$. It is also easy to prove that $T_n^{(m-1)}(z) = \frac{w^{-(m-1)}}{h_1^{m-1}} + O(w^{-m})$. Then, the main contribution follows from the last term of the Taylor expansion of n^{-z}, that is $n(-1)^m \frac{w^m \ln^m n}{m!}$. After some algebra, we can show that

$$l_n^m \sim n \frac{\ln^m n}{h_1^m} \tag{3.36}$$

Noting that $s_n^m = l_n^m/n$ one proves Proposition 2(iii), and this completes our analysis.

APPENDIX. A FORMULA ON AN ALTERNATING SUM: *Alternative Approach to S.O. Rice Method.*

Let us compute

$$F_m(n) = \sum_{k=m}^{n} (-1)^k \binom{n}{k} f(k),$$

where $f(k)$ has an analytical continuation $f(z)$ in a complex plane.

Theorem A. If $f(z)$ is analytical left to the line $(½ - m - i\infty,\ ½ - m - i\infty)$, then

$$\sum_{k=m}^{n} (-1)^k \binom{n}{k} f(k) = \frac{1}{2\pi i} \int_{½ - m - i\infty}^{½ - m + i\infty} \frac{\Gamma(z) f(-z)}{n^{\underline{z}}} dz, \tag{A1}$$

where $n^{\underline{z}} \stackrel{def}{=} \frac{\Gamma(n+1)}{\Gamma(n+1-z)}$, and $\Gamma(z)$ is the gamma function.

Proof. Let us consider the integral, and apply residue theory. In the left half plane the function under the integral is analytical except at points $-k$, $k \geq 0$ is integer, where $\Gamma(z)$ has singularities of value $\frac{(-1)^k}{k!}$. Hence $\left[\int_{(c)} \text{ stands for } \frac{1}{2\pi} \int_{c - i\alpha}^{c + i\infty} \right]$

$$\int_{(\frac{1}{2} - m)} \frac{\Gamma(z) f(-z)}{n^{\underline{z}}} dz = \sum_{k=m}^{\infty} \frac{(-1)^k}{k!} f(-(-k)) n^{\underline{k}}.$$

Note that $n^{\underline{k}} = n(n-1) \ldots (n-1+k) = \frac{n!}{(n-k)!}$, therefore, (A1) is proved.

□

Remark. Some additional conditions on $f(z)$ must be imposed to guarantee (A1). Roughly speaking, the function $f(z)$ cannot grow to fast at infinity. The details can be found in [20].

Let now

$$F_{m,r}(n) = \sum_{k=m}^{n} (-1)^k \binom{n}{k} \binom{k}{r} f(k) \qquad r \textit{ is an integer}$$

Corollary A1. If $f(z)$ is analytical to the left of line $(\frac{1}{2} - [m-r]^+ - i\infty,\ \frac{1}{2} - [m-r]^+ + i\infty)$, then

$$F_{m,r}(n+r) = (-1)^r \binom{n+r}{r} \int_{(\frac{1}{2} - [m-r]^+} \frac{\Gamma(z) f(r-z)}{n^{\underline{z}}} dz.$$

Proof. Using well-known combinatorial identities [23] we find

$$F_{m,r}(n+r) = \sum_{k=m}^{n+r} (-1)^k \binom{n+r}{k} \binom{k}{r} f(k) = \binom{n+r}{r} (-1)^r \sum_{k=[m-r]^+}^{n} (-1)^k \binom{n}{k} f(k+r),$$

and the rest follows from Theorem A.

□

The function $n^{\underline{z}}$ is not very nice to analyze. However, it can be improved. Note that for any a, by Binnet formula [11], [22],

$$\Gamma(n) = n^{n - \frac{1}{2}} e^{-n} (2\pi)^{\frac{1}{2}} e^{\theta/12n}$$
$$\Gamma(n+a) = (n+a)^{n+a-\frac{1}{2}} e^{-n-a} (2\pi)^{\frac{1}{2}} e^{\theta/12(n+a)} \qquad 0 < \theta < 1,$$

hence

$$\frac{\Gamma(n+a)}{\Gamma(n)} = e^{-a} \left(1 + \frac{a}{n} \right)^{n - ½} (n+a)^a \ \exp\,[-\,\theta\ a/1(12\ n(n+a))].$$

But

$$\left(1 + \frac{a}{n} \right)^{n - ½} = e^a\ (n + a^2 O(n^{-1}))$$

$$(n+a)^a = n^a \left(1 + \frac{a}{n} \right)^a = n^a\ (1 + aO(n^{-1}))$$

$$\exp\,[-\,\theta a/12n(n+a)] = 1 + aO(n^{-2}).$$

Finally,

$$\frac{\Gamma(n+a)}{\Gamma(n)} = n^a(1 + a^2O(n^{-1}))(1 + aO(n^{-1}))(1 + aO(n^{-2})) =$$

$$n^a(1 + aO(n^{-1})).$$

It can be shown, in fact, that $\Gamma(n+a) = \Gamma(n) = n^a[1 + \binom{a}{2} O(n^{-1})]$. We have proved

Lemma . For all complex numbers z

$$\frac{1}{n^{\underline{z}}} = \frac{\Gamma(n\ + 1-z)}{\Gamma(n+1)} = n^{-z}\ [1 + zO(n^{-1})].$$

□

So we have obtained

Theorem B. The following holds

$$F_{m,r}(n) = \int_{(+ ½ - m)} \Gamma(z) f(-z) n^{-z} dz\ + e_n$$

where the error function is

$$e_n = O(n^{-1}) \int_{(½ - m)} z\Gamma(z) f(-z) n^{-z}\ dz.$$

□

Corollary B1.

$$F_{m,r}(n+r) = \frac{(-1)^r}{r!} \int_{(½ - [m-r]^+)} \Gamma(z) f(r-z) n^{r-z}\ + e_{n,r}$$

where

$$e_{n,r} = O(n^{-1}) \int_{(\frac{1}{2} - [m-r]^{*})} z\Gamma(z) f(r-z) n^{r-z} \, dz.$$

□

Note that $e_{n,r} = o(n)$.

REFERENCES

[1] Abramowitz, M., Stegun, I., Handbook of mathematical functions, Dover, New York (1964).

[2] Aho, A., Hopcroft, J. and Ullman, J., Data structures and algorithms, Addison-Wesley (1983).

[3] Berger, T., Poisson multiple access for packet broadcast channels, *IEEE Trans. on Information Theory*, IT-30, (1984), 745 - 751.

[4] Frdelyi, A., Higher transcendental functions, McGraw-Hill Book (1953).

[5] Fagin R., Nievergelt, J., Pippenger, N. and Strong H., Extendible hashing: A fast access method for dynamic files, *ACM TODS* 4 (1979), 315–344.

[6] Flajolet, Ph., On the performance evaluation of extendible hashing and trie searching, *Acta Informatica* 20 (1983), 345–369.

[7] Flajolet, Ph. and Sedgewick, R., Digital search trees revisited, *SIAM J. Comput.*, 15 (1986), 748–767.

[8] Flajolet, Ph, Puech, C., Partial match retrieval of multidimensional data, *Journal of ACM*, 33, (1987), 371-407.

[9] Gallager R., Conflict resolution in a random access broadcast networks, *Proc. AFOSR Workshop in Communication Theory and Applications* (1978), 74–76.

[10] Gonnet G., Handbook of algorithms and data structures, Addison-Wesley (1984).

[11] Henrici, P., Applied and computational complex analysis, John Wiley & Sons, New York (1977).

[12] Jacquet, Ph. and Regnier, M., Limiting distributions for trie parameters, in *Lecture Notes in Computer Science*, 214, (1986), pp. 196 - 210.

[13] Kirschenhofer, P. and Prodinger, H., Some further results on digital trees, in *Automata, Language and Machines*, (L. Kott, ed.), Springer Lecture Notes in Computer Science, 226, (1986), pp. 177 - 185.

[14] Knuth, D. The art of computer programming, sorting and searching, Addison-Wesley (1973).

[15] Pittel, B., Paths in a random digital tree: Limiting distributions, *Adv. Appl. Probab.*, 18 (1986), 139–155.

[16] Sedgewick, R., Algorithms, Addison-Wesley (1983).

[17] Szpankowski, W., Some results on V-ary asymmetric tries, to appear in *Journal of Algorithms.*

[18] Szpankowski, W., How much on the average is the Patricia trie better, *Proc. of 24-th Annual Allerton Conference*, (1986), pp. 314 - 323.

[19] Szpankowski, W., Patricia tries again revisited, Purdue University, CSD TR-625, September 1986, submitted to a journal.

[20] Szpankowski, W. On an alternative sum useful in the analysis of some data structure, Purdue University, CSD TR-676, April 1987, (submitted to a journal).

[21] Riordan, J., Combinatorial identities, John Wiley & Sons (1968).

[22] Whittaker, E. and Watson, G., A course of modern analysis, Cambridge Press (1935).