MULTIDIMENSIONAL DIGITAL SEARCHING AND SOME NEW PARAMETERS IN TRIES

April 26, 1996

Peter Kirschenhofer*
Dept. of Algebra&Discrete Math.
Technical University of Vienna
Vienna A-1040
Austria

Helmut Prodinger†
Dept. of Algebra&Discrete Math.
Technical University of Vienna
Vienna A-1040
Austria

Wojciech Szpankowski‡
Dept. of Computer Science
Purdue University
W. Lafayette, IN 47907
U.S.A.

Abstract

Multidimensional digital searching (M-d tries) is analyzed from the view point of partial match retrieval. Our first result extends the analysis of Flajolet and Puech of the average cost of retrieval under the Bernoulli model to biased probabilities of symbols occurrences in a key. The second main finding concerns the variance of the cost of the retrieval in the unbiased case. This variance is of order $O(N^{1-s/M})$ where N is the number of records stored in a M-d trie, and s is the number of specified components in a query of size M. For $M = 2$ and $s = 1$ we present a detailed analysis of the variance, which identifies the constant at $\sqrt{N}$. This analysis, which is the central part of our paper, requires certain series transformation identities which go back to Ramanujan. In the Appendix we provide a Mellin transform approach to these results.

*†This work was supported by Fonds zur Förderung der Wissenschaftlichen Forschung Project P7497-TEC

‡Support for this research was provided in part by NSF Grant INT-8912631, CCR-9201078 and NCR-9206315, in part by AFOSR grant 90-0107, NATO Grant 0057/89, and grant R01 LM05118 from the National Library of Medicine.

1. INTRODUCTION

Multidimensional searching is an important algorithmic concept in modern computer science. In particular, retrieval of multidimensional data found applications in the design of data base systems and graphics. Bentley [6] and Rivest [22] are founding fathers of multidimensional searching, namely digital M-d trees and M-d search trees. An early description of these structures can be found in volume three of Knuth's book [19], and a more detailed discussion is in the book of Mahmoud [20]. Several applications of multidimensional searching are discussed in the paper of Flajolet and Puech [11], which also presents a thorough analysis of partial match retrieval for multidimensional data.

In this paper we concentrate on M-d digital trees (i.e., M-dimensional tries) and partial match retrieval in such a data structure. In general, a *trie* stores a set of, say, N items, which are represented by keys from Σ^∞, the set of all infinite sequences over a finite alphabet Σ. A trie is composed of branching nodes, also called *internal nodes*, and *external nodes* that store the keys. In addition, we assume that every external node is able to store only one key. The branching policy at any level, say k, is based on the k-th symbol of a key. For example, for a binary alphabet $\Sigma = \{0, 1\}$, if the k-th symbol in a key is "0", then we branch-out left in the trie, otherwise we go to the right. This process terminates when for the first time we encounter a different symbol between a key that is currently inserted into the trie and all other keys already in the trie. Then, this new key is stored in a newly generated external node. In other words, the access path from the root to an external node (a leaf of a trie) is the minimal unique prefix of the information contained in this external node. (Cf. also [2], [17], [18], [19], [20]).

The M-d *tries* are built in a similar manner, however, this time a key is an M-dimensional tuple. Let the ith key be $K_i = (K_{i1}, \ldots, K_{iM})$, where $K_{ij} \in \Sigma^\infty$, and $1 \le i \le N$. Now we produce for each $1 \le i \le N$ one sequence $\widetilde{K}_i \in \Sigma^\infty$ by regular shuffling of the components $K_{i1}, \ldots, K_{iM}$, and use these new composite keys $\widetilde{K}_1, \ldots, \widetilde{K}_N$ to construct a regular trie. More precisely, let $K_i = (K_{i1}, \ldots, K_{iM})$ and $K_{i\ell} = (k^1_{i\ell}, k^2_{i\ell}, \ldots)$ where $k^m_{i\ell} \in \Sigma$ for $1 \le \ell \le M$ and $m \ge 1$. Then, the new ith key is $\widetilde{K}_i = (k^1_{i1}k^1_{i2}\cdots k^1_{iM}k^2_{i1}k^2_{i2}\cdots k^2_{iM}\cdots)$. In passing, we note that the statistics of the composite keys are the same as the original keys provided the keys are statistically independent.

A *partial match query* $q = (q_1, \ldots, q_M)$, $q_j \in \Sigma^\infty \cup \{*\}$, asks for all records $(r_1, \ldots, r_M)$ with $r_j = q_j$ for all specified components q_j of the query. For example, $q = (S_1, *, S_2)$ asks for all records with (specified) first component S_1, unspecified second component and (specified) third component S_2. Because of the construction of the M-d tries it is convenient to transform a given query q into a one-dimensional *search sequence* $\sigma = \sigma(q) \in (\Sigma \cup \{*\})^\infty$ by regular shuffling, too. That is, the first symbol of σ comes from the first symbol of the first component of q (if the first component is unspecified $*$, then we put $*$ in σ too), the second symbol of σ originates from the first symbol of the second component of q, and so on modulo M. The search sequence is used to search for external nodes in an M-d trie that satisfy the pattern, where "$*$" means to proceed in all directions. The *cost of a query* is defined as the number of internal nodes visited during such a search.

The cost of a query depends on the search sequence σ. In fact, this sequence specifies a subtree in a trie, and can be considered as a new parameter that characterizes the trie. For example, if a query consists only of specified patterns, then the query defines a specific path (depth) in the trie. This path may be of various shapes depending on the search sequence. Indeed, if $\Sigma = \{0, 1\}$ and $\sigma = (1111\cdots)$, then the query finds the right-most internal nodes, and the cost of the query is the length of such a path; the length of a "zig-zag" path can be computed by evaluating the cost of the following search sequence $\sigma = (0101\cdots)$, etc. On the other hand, if the query consists of only unspecified components $*$, then the cost of the query is equivalent to the size of the underlying trie. Finally, with a mixture of specified and unspecified components, the cost of a query represents the size of a subtree of the trie.

In the following, for simplicity of presentation, we restrict ourselves to the binary alphabet $\Sigma = \{0, 1\}$. We adopt the following *probabilistic model*: Symbols from the alphabet are generated independently, and 0 and 1 occur with probabilities p_0 and $p_1 = 1 - p_0$ respectively (*Bernoulli model*). In [11] Flajolet and Puech – using ingenious techniques from complex analysis and differential equations – presented a very detailed evaluation of the *expected* cost of a query, assuming equal probabilities for symbol occurrences. Observe that in this unbiased instance the expected cost of a fixed query q does not depend on the distribution of the zeroes and ones in the search sequence $\sigma = \sigma(q)$, but only on its *search pattern* $\omega \in \{S, *\}^M$: If we substitute each specified element of σ by S, we obtain a sequence in $\{S, *\}^\infty$ which has period M (since σ was constructed by regular shuffling from q), and we denote its prefix of length M by ω.

In this paper we extend Flajolet and Puech's result into two directions. First, we consider the expected cost in the *biased case*, that is, when probabilities of symbol occurrences are not identical. We investigate two aspects of this problem, namely the *deterministic* instance of *fixed search sequences* σ as well as the average cost for random queries with a *fixed search pattern* ω, where we assume that the specified components of the query satisfy the same Bernoulli statistics as the keys of the trie (*nondeterministic instance*). Our motivation for studying the biased case is mainly to see the impact of (e.g. slightly) different probabilities on the order of the cost. It turns out that this order varies only mildly if p_0 is not too close to either 0 or 1 (compare the table in Section 2). This result matches well with some other analyses for biased trie-like structures by Fayolle, Flajolet and Hofri [10], and Jacquet [14].

More importantly, we also analyze the *variance* of the cost in the unbiased case for $M = 2$, which already requires rather sophisticated mathematical techniques. We shall show that this variance is of order $O(\sqrt{N})$, which implies the convergence *in probability* of the cost to its average value. We also conjecture that for general M the variance is of order $O(N^{1-s/M})$ where s is the number of specified components in a query. We point out that this can be difficult to prove, especially if one needs to identify the constant at $N^{1-s/M}$.

This paper is organized as follows. In the next section we present our main results, and discuss some consequences of them. All proofs are delayed till Section 3. The asymptotic analysis is based on solving some recurrences equation through generating function approach, and obtaining asymptotics of some alternating sums through the usage of Rice's formula (cf. [12]) or the Mellin transform approach (cf. [23]). The main difficulty in this analysis lies in proving that some coefficients in the asymptotics of the variance are equal to zero. These kind of problems were already tackled by us in [17] and [18]. However, this time the difficulty is one step higher. More precisely, during the course of our analysis we require transformation results for $F(x) = \sum_{k\geq 1} e^{-kx}/(1 + e^{-2kx})$ and related functions. It turns out that by a very general formula of Ramanujan (cf. [7]) we have $F(x) = \pi/4x - 1/4 + \pi F(\pi^2/x)/x$ for $x > 0$. In the Appendix we give an outline of a Mellin transform proof for this and similar transformation results.

2. MAIN RESULTS

In this section we present our main results concerning the expected cost of a query. We start with the *nondeterministic case* of random queries, with fixed search pattern $\omega \in \{S, *\}^M$, where we assume that the specified components S of the query are generated in the same manner as the keys stored in the trie, namely 0's and 1's chosen independently with probabilities p_0 resp. $p_1 = 1 - p_0$.

Let $\omega^{(k)}$ denote the cyclic shift to the left by k positions of ω. Then $\omega = \omega^{(M)}$. We also write $\omega' = \omega^{(1)}$. Let $H_N^\omega(z)$ be the probability generating function of the cost of a random query with search pattern ω. Then the following recurrence holds for $N \geq 2$:

$$H_N^\omega(z) = \begin{cases} z\sum_{k=0}^N \binom{N}{k} p_0^k p_1^{N-k} H_k^{\omega'}(z) H_{N-k}^{\omega'}(z) & \omega = *\ldots \\ z\sum_{k=0}^N \binom{N}{k} p_0^k p_1^{N-k} \left(p_0 H_k^{\omega'}(z) + p_1 H_{n-k}^{\omega'}(z)\right) & \omega = S\ldots \end{cases} \tag{1}$$

and $H_0^\omega(z) = H_1^\omega(z) = 1$. (Compare the detailed arguments in [11] for the symmetric case.)

The expected cost l_n^ω is the first derivative of $H_N^\omega(z)$ at $z = 1$. Therefore we have

$$l_N^\omega = 1 + \sum_{k=0}^N \binom{N}{k} p_0^k p_1^{N-k} \{\delta(0, x_1) l_k^{\omega'} + \delta(1, x_1) l_{N-k}^{\omega'}\}\ , \tag{2}$$

$l_0^\omega = l_1^\omega = 0$, where $\omega = (x_1, x_2, \ldots, x_M)$ and

$$\delta(b, x) = \begin{cases} p_b & \text{if } x = S, \\ 1 & \text{if } x = *, \end{cases} \qquad b \in \{0, 1\}. \tag{3}$$

Since $\omega = \omega^{(M)}$, recurrence (2) can be solved by iteration, and the following asymptotic result holds (compare Section 3 for all details).

Theorem 1.1. *Let $\omega = (x_1, x_2, \ldots, x_M) \in \{S, *\}^M$ and denote by s_j the number of specified components S in $(x_1, \ldots, x_j)$, with $s = s_M$. Then the expected cost l_N^ω of a random query q with search pattern ω fulfills for large N*

(i) *if $s = M$ (all components specified)*

$$l_N^\omega = \frac{\log N}{-\left(p_0 \log p_0 + p_1 \log p_1\right)} + O(1)\ , \tag{4}$$

(ii) *if $0 < s < M$*

$$l_N^\omega \sim N^{\kappa_0} \left(\frac{\Gamma(-\kappa_0)(1-\kappa_0) \sum_{j=0}^{M-1} \left(p_0^{1+\kappa_0} + p_1^{1+\kappa_0}\right)^{s_j} \left(p_0^{\kappa_0} + p_1^{\kappa_0}\right)^{j-s_j}}{s \frac{p_0^{1+\kappa_0} \log p_0 + p_1^{1+\kappa_0} \log p_1}{p_0^{1+\kappa_0} + p_1^{1+\kappa_0}} + (M-s) \frac{p_0^{\kappa_0} \log p_0 + p_1^{\kappa_0} \log p_1}{p_0^{\kappa_0} + p_1^{\kappa_0}}} + \xi(N) \right) \tag{5}$$

where $0 < \kappa_0 < 1$ is the unique real solution of the equation

$$\left(p_0^{1+\kappa_0} + p_1^{1+\kappa_0}\right)^s \left(p_0^{\kappa_0} + p_1^{\kappa_0}\right)^{M-s} = 1, \tag{6}$$

and $\xi(N)$ is an oscillating function which is zero if $\frac{\log p_0}{\log p_1}$ is irrational.

Observe that for $s = M$ l_N^ω is the same as the average depth of a randomly selected external node ([19], [20]). The denominator $h = -p_0 \log p_0 - p_1 \log p_1$ is the entropy.

For the reader's convenience we present a table of exponents $\kappa_0 = \kappa_0(p_0, \frac{s}{M})$:

$p \backslash \frac{s}{M}$	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{3}{4}$
$\frac{1}{10}$	0.838	0.779	0.648	0.492	0.398
$\frac{1}{5}$	0.799	0.727	0.574	0.405	0.314
$\frac{1}{4}$	0.784	0.708	0.549	0.380	0.290
$\frac{1}{3}$	0.765	0.685	0.521	0.352	0.266
$\frac{1}{2}$	0.750	0.667	0.500	0.333	0.250

We note that the exponent κ_0 varies substantially with s/M, while the variation of κ_0 with respect to p is rather small.

In the *deterministic case* we ask for the expected cost of a fixed query q with associated query sequence $\sigma = (x_1, x_2, \ldots) \in \{0, 1, *\}^\infty$. This time we have the following equations for the probability generating functions

$$H_N^\sigma(z) = \begin{cases} z \sum_{k=0}^N \binom{N}{k} p_0^k p_1^{N-k} H_k^{\sigma'}(z) H_{N-k}^{\sigma'}(z) & \sigma = *\sigma' \\ z \sum_{k=0}^N \binom{N}{k} p_0^k p_1^{N-k} H_k^{\sigma'}(z) & \sigma = 0\sigma' \\ z \sum_{k=0}^N \binom{N}{k} p_0^k p_1^{N-k} H_{N-k}^{\sigma'}(z) & \sigma = 1\sigma'\ , \end{cases} \tag{7}$$

and for the expected cost

$$l_N^\sigma = 1 + \sum_{k=0}^{N} \binom{N}{k} p_0^k p_1^{N-k} \{\varepsilon(0,x_1) l_k^{\sigma'} + \varepsilon(1,x_1) l_{N-k}^{\sigma'}\} \ , \tag{8}$$

where $\sigma = (x_1, x_2, \ldots)$ and, for $b \in \{0,1\}$, $x \in \{0,1,*\}$

$$\varepsilon_{b,x} = \begin{cases} 1 & \text{if } x = * \\ 1 & \text{if } x \in \{0,1\},\ b = x \\ 0 & \text{if } x \in \{0,1\},\ b \neq x \ . \end{cases} \tag{9}$$

In general, (8) can be solved by iteration, but the asymptotic behavior of the solution will depend heavily on the structure of the 0,1–pattern of σ (observe that the $*$–pattern has period M). In the case where σ is periodic with period k similar analytic techniques as in the nondeterministic case can be applied. Since the structure of the result is quite similar to Theorem 1.1., we only give the exponent of N in case (ii).

Theorem 1.2. *Let $\sigma \in \{0,1,*\}^\infty$ be a fixed search sequence of period k with n_a symbols $a \in \{0,1,*\}$ per period. Then, for large N,*

(i)

$$l_N^\sigma \sim \frac{k \log N}{-n_0 \log p_0 - n_1 \log p_1} + O(1), \qquad \textit{if } n_* = 0, \quad \textit{and} \tag{10}$$

(ii)

$$l_N^\sigma \sim N^{\kappa_0} \cdot (C(\sigma, p_0) + \xi_1(N)) \ , \qquad \textit{if } n_* > 0, \tag{11}$$

where $0 < \kappa_0 < 1$ is the unique solution of

$$p_0^{\kappa_0 n_0} p_1^{\kappa_0 n_1} (p_0^{\kappa_0} + p_1^{\kappa_0})^{n_*} = 1. \tag{12}$$

The next result is our main finding and it concerns the variance of the cost in the case where at least one component of the query is unspecified. This analysis is much more intricate, as demonstrated in our papers [17], [18] that are devoted to the variance of the external path length in a trie and PATRICIA respectively. Therefore, hereafter we assume only the symmetric alphabet, that is, "0" and "1" occur with equal probability 0.5. Observe that under this assumption the nondeterministic cost model coincides with the deterministic one. Now (1) reads

$$H_N^\omega(z) = \begin{cases} z \sum_{k=0}^N \binom{N}{k} 2^{-N} H_k^{\omega'}(z) H_{N-k}^{\omega'}(z) & \omega = *\ldots\ , \\ z \sum_{k=0}^N \binom{N}{k} 2^{-N} H_k^{\omega'}(z) & \omega = S\ldots\ . \end{cases} \tag{13}$$

If the cost of the query ω is denoted by C_N^ω, then the variance var C_N^ω can be evaluated according to the following formula

$$\text{var } C_N^\omega = \frac{d^2}{dz^2} H_N^\omega(z)\bigg|_{z=1} + l_N^\omega - (l_N^\omega)^2, \qquad l_N^\omega = E\, C_N^\omega\ . \tag{14}$$

The evaluation of the second factorial moment of C_N^ω leads to very tedious algebraic manipulations, although the main idea behind it is not too complicated. In the following we restrict our investigations to the case of $M = 2$. As shown in Section 3, the analysis of the centralized moments in this case is a rather sophisticated one.

Theorem 2. (i) *For $\omega = (*, S)$ we have asymptotically*

$$\text{var } C_N^\omega \sim \sqrt{N}\Big(A_1 + \tau_1(\log_2 \sqrt{N})\Big), \qquad A_1 \simeq 2.09184\ , \tag{15}$$

where the constant A_1 can be expressed analytically as

$$\begin{aligned} A_1 \quad &= \quad \frac{7\sqrt{\pi}(1+\sqrt{2})}{2L} - 2\sqrt{\pi}\left(\frac{1}{L}\left[\frac{73+25\sqrt{2}}{16\sqrt{2}} - \frac{4}{3\sqrt{3}}\right]\right. \\ &+ \quad \left.\frac{2}{L}\sum_{\ell\geq 2}\binom{1/2}{\ell}\frac{(\ell-1)(\ell+\frac{1}{2})(1+2^{-\ell})2^{-\ell}}{1-2^{\frac{1}{2}-\ell}}\right)\ , \end{aligned} \tag{16}$$

with $L = \log 2$, and $\tau_1(x)$ is a continuous periodic function of period 1, mean 0 and $|\tau_1(x)| < 5 \cdot 10^{-3}$.
For $\omega = (S, *)$ we have

$$\text{var } C_N^\omega \sim \sqrt{N}\Big(\frac{A_1}{\sqrt{2}} + \tau_2(\log_2 \sqrt{N})\Big), \qquad \frac{A_1}{\sqrt{2}} \simeq 1.47916\ , \tag{17}$$

where $\tau_2(x)$ is again of period 1, mean 0 and $|\tau_2(x)| < 5 \cdot 10^{-3}$.

(ii) *The cost of the query ω converges in probability to $EC_N^\omega = \ell_N^\omega$, that is, the following holds*

$$\lim_{N\to\infty} \Pr\{|\ C_N^\omega / l_N^\omega - 1\ | > \varepsilon\} = 0 \tag{18}$$

for any $\varepsilon > 0$.

Remarks.

(i) *Periodic Fluctuations.* The Fourier coefficients of τ_1 and τ_2 can be computed too, but we omit them here since the expressions are rather complicated. Estimating these coefficients we obtain the upper bounds for the amplitudes.

(ii) *Proof of Theorem 2(ii).* We just note that part (ii) of Theorem 2 follows directly from part (i) and Chebyshev's inequality.

(iii) *Extensions.* Our results can be extended in many directions. For example, it is easy to see that they hold for V-ary symmetric tries, that is, when $\Sigma = \{1, 2, \cdots, V\}$. It may be interesting to extend our Theorem 2 to an asymmetric alphabet, and it is even more challenging to assume some dependency among symbols in the alphabet Σ. Finally, one may replace the regular tries by other digital structures such as Patricia tries and digital search trees. Some preliminary results in this direction are reported in [16]. Also, a rigorous analysis of suffix trees with partial match retrieval might be of some interest.

3. ANALYSIS

In this section we prove Theorems 1.1, 1.2 and 2.

3.1 The Expected Cost in the Asymmetric Case

Let us consider the nondeterministic case at first.

We start our analysis from the recurrence equation (2) for the average cost l_N^ω of a random query with search pattern $\omega = (x_1, \ldots, x_M) \in \{S, *\}^M$. Let $L^\omega(z)$ be the exponential generating function of the cost, that is, $L^\omega(z) = \sum_{N=0}^\infty l_N^\omega z^N/N!$. Then recurrence (2) can be transformed into the following functional equation

$$L^\omega(z) = \delta(0, x_1) L^{\omega'}(zp_0) e^{p_1 z} + \delta(1, x_1) L^{\omega'}(zp_1) e^{p_0 z} + b(z) \ ,$$

where $b(z) = e^z - 1 - z$. Introducing the transform $\widetilde{L}^\omega(z) = e^{-z} L^\omega(z)$ we reduce the above to

$$\widetilde{L}^\omega(z) = \delta(0, x_1) \widetilde{L}^{\omega'}(zp_0) + \delta(1, x_1) \widetilde{L}^{\omega'}(zp_1) + \widetilde{b}(z) \ ,$$

with $\widetilde{b}(z) = e^{-z}b(z)$. Using the fact that $\omega = \omega^{(M)}$, we obtain after M iterations

$$\begin{aligned}\widetilde{L}^{\omega}(z) &= \sum_{\mathbf{b}\in\Theta_M} \delta(b_1, x_1)\cdots\delta(b_M, x_M)\,\widetilde{L}^{\omega}(zp_{b_1}\cdots p_{b_M}) \\ &+ \sum_{j=0}^{M-1}\sum_{\mathbf{b}\in\Theta_j} \delta(b_1, x_1)\cdots\delta(b_j, x_j)\,\widetilde{b}(zp_{b_1}\cdots p_{b_j}) \ , \end{aligned} \tag{19}$$

where $\Theta_j = \{0,1\}^j$. This functional equation is easy to solve. Comparing the coefficients of $\widetilde{L}^{\omega}(z)$ at $z^N/N!$ we get

$$\widetilde{l}_N^{\omega} = \frac{(-1)^N(N-1)\sum_{j=0}^{M-1}\sum_{\mathbf{b}\in\Theta_j}\prod_{i=1}^{j}\delta(b_i,x_i)p_{b_i}^N}{1-\sum_{\mathbf{b}\in\Theta_M}\prod_{i=1}^{M}\delta(b_i,x_i)p_{b_i}^N} \ . \tag{20}$$

From the definition of $\delta(b_i, x_i)$ it follows that

$$\sum_{\mathbf{b}\in\Theta_j}\prod_{i=1}^{j}\delta(b_i,x_i)p_{b_i}^N = \left(p_0^{1+N}+p_1^{1+N}\right)^{s_j}\left(p_0^N+p_1^N\right)^{j-s_j} =: P(N;s_j)$$

where s_j is the number of components S among $(x_1,\ldots,x_j)$. Therefore, with $s = s_M$,

$$\widetilde{l}_N^{\omega} = \frac{(-1)^N(N-1)\sum_{j=0}^{M-1}P(N;s_j)}{1-P(N;s)}, \qquad N \geq 1 \ . \tag{21}$$

To recover the original average cost l_N^{ω} we note that $l_N^{\omega} = \sum_{k=1}^{N}\binom{N}{k}\widetilde{l}_k^{\omega}$.

The asymptotics of l_N^{ω} can now be obtained by a Mellin transform approach (cf. [23]) or Rice's formula (compare Section 3.2). In particular, using [23] we translate l_N^{ω} into the following Mellin-like integral

$$l_N^{\omega} = \frac{1}{2\pi i}\int_{-3/2-i\infty}^{-3/2+i\infty}\frac{\Gamma(z)N^{-z}(z+1)\sum_{j=0}^{M-1}P(-z;s_j)}{1-P(-z;s)}dz\cdot\left(1+O(\frac{1}{N})\right) \ .$$

Shifting the contour of integration to the right and computing the residues of the integrand we finally obtain Theorem 1.1. Note that in the case $s = M$ there is a double pole at $z = 0$ which produces the logarithm function in Theorem 1.1(i). In the case $0 < s < M$ the dominating pole is a simple pole at $z = -\kappa_0$. Additional simple poles occur at the nonreal zeroes of the denominator. It can be shown that nonreal zeroes at $-\kappa_0 + iy_k$ occur if and only if $\frac{\log p_1}{\log p_0}$ is rational (compare the similar reasoning in [14]). They constitute the fluctuating function $\xi(N)$. It is easy to prove that $0 < \kappa_0 < 1$.

In the deterministic case of a fixed query sequence $\sigma = (x_1, x_2, \ldots) \in \{0, 1, *\}^\infty$ we iterate recurrence (8), observe

$$\sum_{b\in\Theta_j} \prod_{i=1}^{j} \varepsilon(b_i, x_i) p_{b_i}^N = p_0^{Nn_0(\sigma_j)} p_1^{Nn_1(\sigma_j)} \left(p_0^N + p_1^N\right)^{n_*(\sigma_j)} =: Q(N, \sigma_j).$$

where $\sigma_j = (x_1, x_2, \ldots, x_j)$ and get the solution

$$\widetilde{l}_N^\sigma = (-1)^N (N-1) \sum_{j\geq 0} Q(N, \sigma_j). \tag{22}$$

For general σ a suitable analytic continuation of $\sum_{j\geq 0} Q(z, \sigma_j)$ will not exist, but there is one for instance if σ has period k. Then

$$\sum_{j\geq 0} Q(z, \sigma_j) = \frac{\sum_{j=0}^{k-1} Q(z, \sigma_j)}{1 - Q(z, \sigma_k)}$$

is meromorphic and we obtain Theorem 1.2 by the same method as Theorem 1.1.

3.2 The Analysis of the Variance in the Symmetric Case

We first consider $\omega = (*, S)$, $S \in \{0, 1\}$. Before attacking the second factorial moment of the cost C_N^ω we perform a more precise evaluation of the average cost l_N^ω.

From (21) we have

$$l_N^\omega = \sum_{k=1}^{N} \binom{N}{k} (-1)^k f(k) \quad \text{with} \quad f(z) = \frac{(z-1)(1+2^{1-z})}{1-2^{1-2z}}\ . \tag{23}$$

The alternating sum can be treated by the following lemma.

Lemma. (Rice's method, cf. [12], [17], [18]) *Let $\mathcal{C}$ be a path surrounding the points $\{j, j+1, \ldots, N\}$, and $f(z)$ be an analytical continuation of $f(k)$ inside $\mathcal{C}$. Then*

$$\sum_{k=j}^{N} \binom{N}{k} (-1)^k f(k) = -\frac{1}{2\pi i} \int_{\mathcal{C}} [N; z] f(z) dz\ , \tag{24}$$

where

$$[N; z] = \frac{\Gamma(-z)\Gamma(N+1)}{\Gamma(N+1-z)} = \frac{(-1)^{N-1} N!}{z(z-1)\cdots(z-N)}\ . \tag{25}$$

Extending the contour of the integration it can be shown that under suitable growth conditions for $f(z)$ the alternating sum in (24) equals $\sum \mathrm{Res}([N; z] f(z)) +$

$O(N^c)$, where the sum is taken over all residues with real part greater than c that are different from $j, \ldots, N$.

The function $f(z)$ in (23) has single poles at $z_k = \frac{1}{2} + \frac{1}{2}\chi_k$ where $\chi_k = 2\pi i k/L$ and $L = \log 2$. Computing the residues at z_k and additionally at $z = 0$ (coming from the function $[N; z]$), we finally obtain

$$l_N^\omega = \sqrt{N}\left(\sqrt{\pi}\frac{1+\sqrt{2}}{2L} + \tau^\omega\left(\log_2\sqrt{N}\right)\right) - 3 + o(1), \tag{26}$$

where the fluctuating function $\tau^\omega(x) = \sum_{k\neq 0}\tau_k^\omega e^{2k\pi i x}$ has the Fourier coefficients

$$\tau_k^\omega = \frac{1}{2L}\left(1+\sqrt{2}(-1)^k\right)\Gamma\left(\frac{-1-\chi_k}{2}\right)\left(\frac{-1+\chi_k}{2}\right) . \tag{27}$$

Now we deal with the second factorial moment. Let $w_N^\omega = \left(H_N^\omega\right)''(1)$, and $W^\omega(z)$ be its exponential generating function. From now on we define $\widetilde{A}(z) = A(z)e^{-z}$ for any function $A(z)$ under consideration. From (13), after differentiating it twice, we obtain

$$\begin{aligned}
\widetilde{W}^\omega(z) &= 2\widetilde{W}^{\omega\prime}\left(\frac{z}{2}\right) + 4\widetilde{L}^{\omega\prime}\left(\frac{z}{2}\right) + 2\left(\widetilde{L}^{\omega\prime}\left(\frac{z}{2}\right)\right)^2\\
\widetilde{W}^{\omega\prime}(z) &= \widetilde{W}^\omega\left(\frac{z}{2}\right) + 2\widetilde{L}^\omega\left(\frac{z}{2}\right) .
\end{aligned}$$

For simplicity of presentation, $\widetilde{W}^\omega(z) = \widetilde{R}^\omega(z) + \widetilde{T}^\omega(z)$ where

$$\widetilde{R}^\omega(z) = 2\widetilde{R}^{\omega\prime}\left(\frac{z}{2}\right) + 4\widetilde{L}^{\omega\prime}\left(\frac{z}{2}\right), \qquad \widetilde{R}^{\omega\prime}(z) = \widetilde{R}^\omega\left(\frac{z}{2}\right) + 2\widetilde{L}^\omega\left(\frac{z}{2}\right) \tag{28}$$

and

$$\widetilde{T}^\omega(z) = 2\widetilde{T}^{\omega\prime}\left(\frac{z}{2}\right) + 2\left(\widetilde{L}^{\omega\prime}\left(\frac{z}{2}\right)\right)^2, \qquad \widetilde{T}^{\omega\prime}(z) = \widetilde{T}^\omega\left(\frac{z}{2}\right) . \tag{29}$$

The solution of the first system of recurrences, namely (28), is easy. Note that it reduces to the following functional equation

$$\widetilde{R}^\omega(z) = 2\widetilde{R}^\omega\left(\frac{z}{4}\right) + 4\widetilde{L}^{\omega\prime}\left(\frac{z}{2}\right) + 4\widetilde{L}^\omega\left(\frac{z}{4}\right) .$$

Comparing coefficients we immediately find that

$$\widetilde{r}_N^\omega = \frac{(-1)^N(N-1)2^{2-N}}{\left(1-2^{1-2N}\right)^2}\left[1+2^{1-N}+2^{1-2N}\right] . \tag{30}$$

Now we deal with (29), which is more intricate. We need a formula for the coefficients of $\widetilde{M}^{\omega\prime}(z) = \left(\widetilde{L}^{\omega\prime}(z)\right)^2$ that avoids the convolution originating from the square. But from (19)

$$\widetilde{L}^{\omega\prime}(z) = 2\widetilde{L}^{\omega\prime}\left(\frac{z}{4}\right) + \widetilde{b}(z) + \widetilde{b}\left(\frac{z}{2}\right) .$$

Taking it to the power two, we obtain

$$\widetilde{M}^{\omega\prime}(z) = 4\widetilde{M}^{\omega\prime}\left(\frac{z}{4}\right) + 4\left(\widetilde{b}(z) + \widetilde{b}\left(\frac{z}{2}\right)\right)\widetilde{L}^{\omega\prime}\left(\frac{z}{4}\right) + \left(\widetilde{b}(z) + \widetilde{b}\left(\frac{z}{2}\right)\right)^2 ,$$

which allows to express the coefficients of $\widetilde{M}^{\omega\prime}(z)$ in the desired manner.

Putting everything together we find after some tedious algebra that $w_N^\omega = \sum_{k=2}^N \binom{N}{k}(-1)^k f(k)$ with $f(z)$ as below

$$\begin{aligned} f(z) &= \frac{(z-1)2^{1-z}}{(1-2^{1-2z})(1-2^{2-2z})}\left[5 + 2^{3-z} + 2^{3-2z} + z2^{z-2} - 2^z + \frac{z}{4} + \left(z - \frac{9}{2}\right)\left(\frac{3}{2}\right)^{z-2}\right] \\ &+ \frac{2^{3-2z}}{(1-2^{1-2z})(1-2^{2-2z})}\sum_{l\geq 2}\binom{z}{l}\frac{(l-1)(z-l-1)\left(1+2^{-l}\right)\left(1+2^{z-l}\right)2^{-l}}{1-2^{1-2l}} . \end{aligned} \tag{31}$$

Now we can apply our Lemma to the alternating sum w_N^ω to obtain an asymptotic expansion. Note that there are two sets of poles that contribute to the leading terms of the asymptotics. Namely, $z_{k,1} = 1 + \frac{1}{2}\chi_k$ and $z_{k,2} = \frac{1}{2} + \frac{1}{2}\chi_k$ where $k = 0, \pm 1, \cdots$. Easy computations reveal finally that

$$\begin{aligned} w_N^\omega &\sim \frac{N}{L}\left\{8 - \frac{7}{12} - 2\sum_{l\geq 2}(-1)^l \frac{\left(1+2^{-l}\right)\left(1+2^{1-l}\right)2^{-l}}{1-2^{1-2l}} + \tau_3(\log_2\sqrt{N})\right\} \\ &- 2\sqrt{\pi N}\left\{\frac{1}{L}\left[\frac{73}{2\sqrt{2}} + \frac{25}{16} - \frac{4}{3\sqrt{3}}\right]\right. \\ &+ \left.\frac{2}{L}\sum_{l\geq 2}\binom{1/2}{l}\frac{(l-1)\left(l+\frac{1}{2}\right)\left(1+2^{-l}\right)2^{-l}}{1-2^{\frac{1}{2}-l}} + \tau_4(\log_2\sqrt{N})\right\} \end{aligned} \tag{32}$$

with periodic fluctuations $\tau_3(x)$ and $\tau_4(x)$ of mean (i.e. zeroeth Fourier coefficient) zero.

Now we are ready to put everything together. In particular, we note that, *apart from continuous fluctuations of mean zero*,

$$\operatorname{var} C_N^\omega = w_N^\omega + l_N^\omega - (l_N^\omega)^2 = BN + A_1\sqrt{N} + o(\sqrt{N}) , \tag{33}$$

where A_1 and B are constants. Note that

$$B = \frac{1}{L}\left\{8 - \frac{7}{12} - 2\sum_{l\geq 2}(-1)^l \frac{\left(1+2^{-l}\right)\left(1+2^{1-l}\right)2^{-l}}{1-2^{1-2l}}\right\} - \pi\frac{\left(1+\sqrt{2}\right)^2}{4L^2} - [(\tau^\omega)^2]_0 \ ,$$

where $[(\tau^\omega)^2]_0$ is the mean of the square of the periodic function $\tau^\omega(x)$ defined in (26). In the lines below we prove that $B = 0$. As in [17] and [18], we point out that the term $[(\tau^\omega)^2]_0$ significantly contributes to the cancellation of terms in B.

Considering the terms of order N, after having proved $B = 0$, the remaining fluctuating function has to be identically equal to zero since otherwise the variance would be negative for infinitely many values of N, which is impossible. For more detailed arguments the reader is referred to our papers [17] and [18].

To find a transformation for $[(\tau^\omega)^2]_0$ we proceed as follows. Note that

$$[(\tau^\omega)^2]_0 = 2\sum_{k\geq 1}\tau_k^\omega\tau_{-k}^\omega = \frac{2}{4L^2}\sum_{k\geq 1}\left(3+2\sqrt{2}(-1)^k\right)\Gamma\left(\frac{1-\chi_k}{2}\right)\Gamma\left(\frac{1+\chi_k}{2}\right) \ .$$

Now we use the formula (cf. [1]) $\Gamma(z)\Gamma(1-z) = \pi/\sin\pi z$ and obtain

$$\Gamma(\frac{1-\chi_k}{2})\Gamma(\frac{1+\chi_k}{2}) = \frac{\pi}{\sin(\pi/2+ik\pi^2/L)} = \frac{\pi}{\cos(ik\pi^2/L)} = \frac{\pi}{\cosh(k\pi^2/L)} = 2\pi\frac{e^{-k\pi^2/L}}{1+e^{-2k\pi^2/L}} \ ,$$

so that

$$[(\tau^\omega)^2]_0 = \frac{\pi}{L^2}\sum_{k\geq 1}\left(3+2\sqrt{2}(-1)^k\right)\frac{e^{-k\pi^2/L}}{1+e^{-2k\pi^2/L}} \ . \tag{34}$$

Let us define two new functions

$$F(x) = \sum_{k\geq 1}\frac{e^{-kx}}{1+e^{-2kx}} \qquad \text{and} \qquad G(x) = \sum_{k\geq 1}\frac{(-1)^{k-1}e^{-kx}}{1+e^{-2kx}} \ . \tag{35}$$

Then, (34) in terms of $F(x)$ and $G(x)$ becomes

$$[(\tau^\omega)^2]_0 = \frac{3\pi}{L^2}F\left(\frac{\pi^2}{L}\right) - \frac{2\sqrt{2}\pi}{L^2}G\left(\frac{\pi^2}{L}\right) \ . \tag{36}$$

In order to prove $B = 0$, we use a series transformation for $F(x)$ and $G(x)$. We start with

$$F(x) = \sum_{j\geq 0}(-1)^j\sum_{k\geq 1}e^{-k(2j+1)x} = \sum_{j\geq 0}\chi(j)\frac{1}{e^{jx}-1}$$

where

$$\chi(j) = \begin{cases} 0 & \text{for } j \text{ even} \\ 1 & \text{for } j \equiv 1 \bmod 4 \\ -1 & \text{for } j \equiv 3 \bmod 4 \end{cases}$$

In [7] on page 258 we find the following Entry 11 from *Ramanujan's Notebooks.*

For α, $\beta > 0$ such that $\alpha\beta = \pi$ and $y \in \mathbb{R}$ with $|y| < \beta/2$ the following holds

$$\alpha \left\{ \frac{1}{4} \sec(\alpha y) + \sum_{k=1}^{\infty} \chi(k) \frac{\cos(\alpha y k)}{e^{k\alpha^2} - 1} \right\} = \beta \left\{ \frac{1}{4} + \frac{1}{2} \sum_{k=1}^{\infty} \frac{\cosh(2\beta y k)}{\cosh(k\beta^2)} \right\} . \tag{37}$$

For $y = 0$ this reads

$$\alpha \left\{ \frac{1}{4} + \sum_{k=1}^{\infty} \chi(k) \frac{1}{e^{k\alpha^2} - 1} \right\} = \beta \left\{ \frac{1}{4} + \frac{1}{2} \sum_{k=1}^{\infty} \frac{1}{\cosh(k\beta^2)} \right\} .$$

Replacing in the above $\cosh(x)$ by the exponential functions, expanding the geometric series and rearranging the sums we obtain

$$\frac{1}{2} \sum_{k \geq 1} \frac{1}{\cosh(k\beta^2)} = \sum_{j \geq 1} \chi(j) \frac{1}{e^{j\beta^2} - 1} ,$$

so that, taking into account the constraint $\alpha\beta = \pi$, we find

$$\alpha \left(\frac{1}{4} + F(\alpha^2) \right) = \beta \left(\frac{1}{4} + F(\beta^2) \right) .$$

Now, we substitute in the above $\alpha = \sqrt{x}$ and $\beta = \pi/\sqrt{x}$. Then,

$$F(x) = \frac{\pi}{4x} - \frac{1}{4} + \frac{\pi}{x} F\left(\frac{\pi^2}{x} \right) \tag{38}$$

for $x > 0$, which is the desired transformation result. For $G(x)$ we observe that $G(x) = F(x) - 2F(2x)$, hence

$$G(x) = \frac{1}{4} + \frac{\pi}{x} F\left(\frac{\pi^2}{x} \right) - \frac{\pi}{x} F\left(\frac{\pi^2}{2x} \right) . \tag{39}$$

Applying the above to (36) we finally obtain

$$[(\tau^{\omega})^2]_0 = \frac{3}{4L} - \frac{\pi}{4L^2} \left(3 + 2\sqrt{2} \right) + \frac{3 - 2\sqrt{2}}{L} F(L) + \frac{2\sqrt{2}}{L} F\left(\frac{L}{2} \right) . \tag{40}$$

Using (40) $B = 0$ follows by some obvious simplifications. Thus, we have proved $\mathrm{var}\, C_N^\omega = O(\sqrt{N})$. The coefficient A_1 at $\sqrt{N}$ in (33) is given in Theorem 2(i) equation (16).

The second pattern $\omega = (S, *)$ can be analyzed in a similar manner as above.

We finally mention, that it is not at all obvious how to extend this result to the case $M \geq 3$. The reader should note that the key arguments in our proof of Theorem 2(i) are the transformations result (38) and (39). For $M \geq 3$ there is no comparable result in *Ramanujan's Notebooks*, and the alternative proof presented in the Appendix relies heavily on the duplication formula for the gamma function, which is not suitable for $M \geq 3$.

APPENDIX: Mellin Transform Approach to Ramanujan-Like Series Identities

Most proofs of results like Ramanujan's formula (37) resp. (35) use the theory of *modular functions* (cf. [7]). In this appendix we outline the alternative approach using *Mellin integrals*, a line of attack, suggested to us by P. Flajolet, that belongs to the toolkit of analytic number theory (cf. [4]).

To prove (38) we proceed as follows. Let

$$\beta(s) = \sum_{j \geq 0} (-1)^j \frac{1}{(2j+1)^s} .$$

Using definition (35) of $F(x)$, we have

$$F(x) = \sum_{k \geq 1} \frac{e^{-kx}}{1 + e^{-2kx}} = \sum_{j \geq 0} (-1)^j \sum_{k \geq 1} e^{-k(2j+1)x} ,$$

so that the Mellin transform $F^*(s) = \int_0^\infty F(x) x^{s-1} dx$ of $F(x)$ (cf. [9]) becomes

$$F^*(s) = \Gamma(s)\zeta(s)\beta(s) ,$$

where $\zeta(s)$ is Riemann's zeta function [1]. By the Mellin inversion formula this yields

$$F(x) = \frac{1}{2\pi i} \int_{3/2 - i\infty}^{3/2 + i\infty} \Gamma(s)\zeta(s)\beta(s) x^{-s} ds . \tag{41}$$

Now we take the two residues $s = 1$ and $s = 0$ out from the above integral (observe that $\beta(0) = 1/2$ and $\beta(1) = \pi/4$ (cf. [1]) and apply the the duplication formula for

$\Gamma(s)$ to obtain

$$F(x) = \frac{\pi}{4x} - \frac{1}{4} + \frac{1}{2\pi i} \int_{-\frac{1}{2}-i\infty}^{-\frac{1}{2}+i\infty} \frac{1}{\sqrt{\pi}} 2^{s-1} \Gamma\left(\frac{s}{2}\right) \Gamma\left(\frac{s+1}{2}\right) x^{-s} \zeta(s)\beta(s) ds \ .$$

(Observe that the exponential smallness of the Γ-function along vertical lines justifies the shifting of the line integral.)

We now use the functional equations for $\zeta(s)$ and $\beta(s)$, namely

$$\Gamma\left(\frac{s}{2}\right) \zeta(s) = \pi^{s-\frac{1}{2}} \Gamma\left(\frac{1-s}{2}\right) \zeta(1-s) \tag{42}$$

and

$$\beta(1-s)\Gamma\left(1-\frac{s}{2}\right) = 2^{2s-1} \pi^{-s+\frac{1}{2}} \Gamma\left(\frac{s+1}{2}\right) \beta(s) \ . \tag{43}$$

Identity (42) is Riemann's functional equation for $\zeta(s)$, and (43) is an immediate consequence of the functional equation for Hurwitz's ζ-function $\zeta(s,a)$ (cf. [3]), and the fact that

$$\beta(s) = 4^{-s} \left[\zeta(s, \frac{1}{4}) - \zeta(s, \frac{3}{4}) \right] \ .$$

Using (42) and (43), and substituting $1-s=u$, we get

$$F(x) = \frac{\pi}{4x} - \frac{1}{4} + \frac{1}{2\pi i} \int_{\frac{3}{2}-i\infty}^{\frac{3}{2}+i\infty} \pi^{1-2u} \Gamma(u) x^{u-1} \zeta(u) \beta(u) du \ .$$

This proves (38).

Using the above scheme *several other identities* can be proved. Below we just list some of them. Define

$$f(x) = \sum_{k\geq 1} \frac{e^{kx}}{(e^{kx}-1)^2} = \sum_{k\geq 1} \frac{k}{e^{kx}-1}, \tag{44}$$

$$g(x) = \sum_{k\geq 1} \frac{(-1)^{k-1}}{k\,(e^{kx}-1)}, \tag{45}$$

$$h(x) = \sum_{k\geq 1} \frac{1}{k\,(e^{2k\pi x}-1)} \ . \tag{46}$$

Then, the following identities hold:

1. $$1 + 2\sum_{n\geq 1} e^{-\pi n^2 x} = \frac{1}{\sqrt{x}} + \frac{2}{\sqrt{x}} \sum_{n\geq 1} e^{-\pi n^2/x} \qquad \text{Jacobi's } \vartheta\text{–function,}$$

2.

$$f(x) = \frac{1}{x^2}\frac{\pi^2}{6} - \frac{1}{2x} + \frac{1}{24} - \frac{4\pi^2}{x^2} f\left(\frac{4\pi^2}{x}\right) \qquad \text{Ramanujan [7],}$$

3.

$$g(x) = \frac{\pi^2}{12x} - \frac{\log 2}{2} + \frac{x}{24} - g\left(\frac{2\pi^2}{x}\right) \qquad \text{Ramanujan [7],}$$

4.

$$h(x) = \frac{\pi}{12}\left(\frac{1}{x} - x\right) + \frac{\log x}{2} + h\left(\frac{1}{x}\right) \qquad \text{Dedekind and Ramanujan [7].}$$

The latter three identities play a prominent role in the analysis of the variance for some parameters of tries (cf. [15], [17], [18]).

Remark. It is fairly well known that the functional equations for the Riemann ζ-function and the ϑ-function can be derived from each other via the Mellin transform (see e.g., [5], [8]). Here we would like to point out the analogous ϑ-like identity that is related to the functional equation for $\beta(s)$. Let

$$H(x) = \sum_{j\geq 0} (-1)^j (2j+1) e^{-(2j+1)^2 x}\ , \qquad \text{then} \qquad H(x) = \frac{1}{8}\left(\frac{\pi}{x}\right)^{3/2} H\left(\frac{\pi^2}{16x}\right)\ .$$

ACKNOWLEDGMENTS

We would like to thank Philippe Flajolet, INRIA, for his long standing support of our work, and several fruitful discussions on the subject of this paper, as well as the anonymous referees for several helpful remarks.

References

[1] M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions*, Dover, New York 1970.

[2] A.V. Aho, J.E. Hopcroft and J.D. Ullman, *The Design and Analysis of Computer Algorithms,* Addison-Wesley, Reading Mass. 1974.

[3] T. Apostol, *Introduction to Analytical Number Theory*, Springer-Verlag, New York 1976.

[4] T. Apostol, *Modular Functions and Dirichlet Series in Number Theory*, Springer-Verlag, New York 1978.

[5] R. Bellman, *Analytic Number Theory – An Introduction*, Benjamin-Cummings, London 1980.

[6] J. L. Bentley, Multidimensional Binary Search Trees Used for Associative Searching, *Communication of the ACM*, 18, 509-517, 1975.

[7] B.C. Berndt, *Ramanujan's Notebooks. Part II*, Springer-Verlag, New York 1989.

[8] J. Borwein and P. Borwein, *Pi and the AGM*, John Wiley & Sons, New York 1987.

[9] B. Davis, *Integral Transforms and Their Applications*, Springer-Verlag, New York 1978.

[10] G.Fayolle, P.Flajolet and M.Hofri, On a functional equation arising in the analysis of a protocol for a multiaccess broadcast channel, *Advances in Applied Probability*, 18, 1986, 441–472.

[11] P. Flajolet and C. Puech, Partial Match Retrieval of Multidimensional Data, *J. of the ACM*, 33, 371-407, 1986.

[12] P. Flajolet and R. Sedgewick, Digital Search Trees Revisited, *SIAM J. Computing*, 15, 748-767, 1986.

[13] P. Henrici, *Applied and Computational Complex Analysis*, John Wiley&Sons, New York 1977.

[14] P. Jacquet, *Contribution de l'Analyse d'Algorithmes à l'Évaluation des Protocoles de Communications*, Thèse, Université de Paris Sud-Orsay, 1989.

[15] P. Kirschenhofer and H. Prodinger, On Some Applications of Formulae of Ramanujan in the Analysis of Algorithms, *Mathematika*, 38, 14-33, 1991.

[16] P. Kirschenhofer and H. Prodinger, Multidimensional Digital Searching – Alternative Data Structures, *Proc. Random Graphs'91*, John Wiley & Sons, 1992.

[17] P. Kirschenhofer, H. Prodinger and W. Szpankowski, On the Variance of the External Path length in a Symmetric Digital Tries, *Discrete Applied Mathematics*, 25, 129-143, 1989.

[18] P. Kirschenhofer, H. Prodinger and W. Szpankowski, On the balance Property of Patricia Tries: External Path Length Viewpoint, *Theoretical Computer Science*, 68, 1-17, 1989.

[19] D.E. Knuth, *The Art of Computer Programming. Vol. 3*, Addison-Wesley, Reading MA 1973.

[20] H. Mahmoud, *Evolution of Random Search Trees*, John Wiley & Sons, New York, 1992.

[21] M. Régnier and P. Jacquet, New Results on the Size of Tries, *IEEE Trans. Information Theory*, 35, 203-205 (1989).

[22] R. L. Rivest, Partial-Match Retrieval Algorithm, *SIAM J. Computing*, 5, 19-50, 1976.

[23] W. Szpankowski, The Evaluation of an Alternating Sum with Applications to the Analysis of Some Data Structures, *Information Processing Letters*, 28, 13-19, 1988.