1989

Analysis of Digital Tries with Markovian Dependency

P. Jacquet

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

Report Number:
89-906

ANALYSIS OF DIGITAL TRIES
WITH MARKOVIAN DEPENDENCY

P. Jacquet
W. Szpankowski

CSD-TR-906
September 1989

ANALYSIS OF DIGITAL TRIES WITH MARKOVIAN DEPENDENCY

P. Jacquet†
INRIA
Domaine de Volucceau Rocquencourt
78153 Le Chesnay
FRANCE

W. Szpankowski*
Department of Computer Science
Purdue University
W. Lafayette, IN 47907
USA

Abstract

Digital tries occur in a variety of computer and communication algorithms including symbolic manipulations, compiling, comparison based searching and sorting, digital retrieval techniques, algorithms on strings, file systems, codes and communication protocols. It is crucial for all of these applications to design and build robust models of the underlying digital trees. In this paper, we present a complete characterization of tries from the depth viewpoint in a Markovian framework, that is, under the assumption that symbols in a key are Markov-dependent. Our main findings show that asymptotically (i.e., for the number of keys n tending to infinity) the average depth $ED_n \sim 1/h_1 \cdot \log n + c'$, the variance $var\, D_n \sim \alpha \log n + c''$, where h_1 is the entropy of the (Markovian dependent) alphabet, and α is a parameter that becomes zero for *symmetric independent model*, that is, in this case $var\, D_n = O(1)$. We also derive limiting distribution for the depth D_n, and in particular, we show that D_n tends to the *normal distribution* in all cases except the symmetric independent model. This finding implies that in general the m-th moment $E\{(D_n - ED_n)/\sqrt{varD_n}\}^m \to m!/[2^{m/2}(m/2)!]$ for m even, and it tends to zero for m odd. These results extend all previous analyses since most of them have been limited to independent models.

---

† This research is partially supported by NATO Collaborative grant 00571/89.
* This research was primary done while the author was visiting INRIA in Rocquencourt, France. Support was provided in part by NATO Collaborative grant 00571/89, in part by NSF grants NCR-8702115 and CCR-8900305, and in part by grant R01 LM05118 from the National Library of Medicine.

1. INTRODUCTION

Digital trees appear in a variety of computer and communication applications including searching, sorting, dynamic hashing [1,2,3], codes and most recently in tree and stack communication protocols [4–7]. Of particular interest is a digital tree called (radix) *trie*. A trie is a data structure associated with a set $\mathcal{S}$ of (possible) infinite strings built over a finite alphabet $\mathcal{A} = \{\omega_1, \omega_2, \ldots, \omega_V\}$, where ω_i is the i-th letter (symbol) of the V-ary alphabet $\mathcal{A}$. In the most common analytical model of tries, it is assumed that cardinality $|\mathcal{S}|$ of $\mathcal{S}$ is fixed and equal to n, that is, the underlying trees store n records. A trie over $\mathcal{S}$ is built recursively as follows. If $|\mathcal{S}| \le 1$, then $trie(\mathcal{S})$ consists of a single leaf containing this element. If $|\mathcal{S}| = n > 1$, then $\mathcal{S}$ is split into V subsets $\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_V$ where $\mathcal{S}_j$ contains all strings that starts with ω_j symbol, and the $trie(\mathcal{S})$ is obtained by appending $trie(\mathcal{S}_1), \ldots, trie(\mathcal{S}_V)$ as subtrees of the common root. Our interest lies in evaluating the depth of a (randomly) chosen string (key) in a trie that does *not* necessarily assume independence between symbols in the alphabet $\mathcal{S}$. The analysis of tries depends upon assumed probabilistic framework. Any probabilistic model has to address at least three issues, namely dependency among symbols of the alphabet $\mathcal{S}$, dependency among strings, and finally the distribution of $|\mathcal{S}|$, that is, whether $|\mathcal{S}|$ is fixed and deterministic or random.

In this paper we concentrate on the so called *Markovian model* which can be characterized by the following three assumptions:

A1. Symbols of the alphabet $\mathcal{A} = \{\omega_1, \omega_2, \ldots, \omega_V\}$ are Markovian dependent, that is, for a word $x = x_1 x_2 x_3 \ldots$ the probability $p_{ij} = Pr\{x_{t+1} = \omega_j \mid x_t = \omega_i\}$ prescribes the conditional probability of sampling symbol ω_j following symbol ω_i.

A2. Strings of $\mathcal{S}$ are statistically independent.

A3. The number of strings (i.e., the cardinality $|\mathcal{S}|$ of $\mathcal{S}$) is fixed and equal to n.

We note that for the so called *independent model* (no dependency between symbols), the transition probability p_{ij} does not depend on i, that is, $p_{ij} = p_j = Pr\{x_t = \omega_j\}$ for every i. Thus, the Markovian model is a natural extension of the independent model. We also note that the fixed cardinality assumption A3 defines the so called *Bernoulli model* which is postulated throughout the entire paper. Finally we point out that our methodology allows us to extend assumption A1 to include r-dependent stationary Markov chains, however, for simplicity of further presentation, we only deal with assumption A1 (see Remark (ii)).

Our main interest lies in evaluating the depth of a leaf (also called depth of insertion and successful search time) in the Markovian model. This quantity is very useful in estimating complexity of some algorithms, most notably the searching time in any search algorithm based on tries [2], and the length of conflict resolution session for tree-based communication protocols [6,7]. In this paper we prove that the average value ED_n of the depth D_n for large n becomes $ED_n \sim 1/h_1 \cdot \log n + c'$ where c' is a constant and $h_1 = \sum_{i,j} \pi_i p_{ij} \log p_{ij}$ is the entropy of the alphabet $\mathcal{A}$. The variance $var\, D_n$ appears to be $\alpha \cdot \log n + c''$ with $\alpha = 0$ for the *independent symmetric* model, that is, when for all i and j, $p_{ij} = 1/V$. In addition, we present exact distribution function for the depth D_n, and show that D_n tends to the normal distribution in all cases except the independent symmetric model.

Tries have been analyzed in the past by several authors, but most of the analyses confined to independent model [8–13] except the paper of Pittel [14] (convergence in probability) and Régnier [15] (size of tries). Knuth [2], Flajolet [8], and Kirschenhofer and Prodinger [12] analyzed only binary symmetric independent model for tries, and the authors of [12] have additionally obtained the variance of the depth. This has been independently extended by Jacquet and Régnier [9,10] (limiting distribution), Pittel [13] (limiting distribution), and Szpankowski [11] (all moments) to an asymmetric independent model. Authors of [8–12] use extensively complex analysis to derive their results. Finally, Pittel [14] proved convergence *in probability* of the depth D_n for a general dependency between symbols, and Régnier [15] obtained the first leading factor in a Markovian analysis of the *average* size of tries. Our methodology is new, and it is based only on the *inclusion-exclusion rule* which is general enough to be also applied to many other models, most notably a dependent model with correlated strings (relaxation of assumption A2).

This paper is organized as follows. The next section defines our basic model and presents main results. The last section contains proofs.

2. MAIN RESULTS

In this section we present a general idea of our novel approach to evaluate depth for tries. We illustrate the power of this approach by re-deriving depth for independent models. Finally, we present our main results for the Markovian model, delaying all proofs to Section 3, and discuss some consequences of our findings.

Let us start with some definitions that illustrate our new approach to the problem under consideration. In short, we introduce a notion of *alignment* (cf. [16]) that is used throughout the entire paper. We note also that at this moment we to *not* make any assumption regarding distributions of symbols and/or strings, except we adopt for simplicity assumption A3. Let X_1, $X_2, \ldots, X_n \in \mathcal{S}$ be n strings (keys) stored in a trie. We define the alignment C_{ij} between the i-th and the j-th keys, $i \neq j$, as the length of the longest prefix of both X_i and X_j. Thus, $C_{ij} = k$ iff X_i and X_j agree exactly on their first k symbols but differ on their $(k+1)$-st. In [16] we have shown how to express some parameters of tries through the notion of alignments C_{ij}. Here we concentrate on the depth which, as we shall indicate below, is the most difficult to analyze.

Let $D_n(i)$ denote the depth of the i-th key in a trie, that is, the number of nodes in a path from the root to the i-th key. Then, by the definition of the alignment, one immediately finds that

$$D_n(i) = \max \{C_{i1}, \ldots, C_{in}\}$$

Now, let us assume that we randomly select a key, and for such a key we evaluate the depth. Such a depth is called *the average depth*, and we denote it as D_n. At this point, we additionally adopt assumption A2 (independent keys), and then it is easy to see that all $D_n(i)$ are equidistributed, which implies that D_n has the same distribution as $D_n(i)$ for arbitrary i [14]. Therefore, the original definition of D_n reduces to

$$D_n = \max \{C_{12}, C_{13}, \ldots, C_{1n}\} \tag{2.1}$$

The above formula on D_n is used to derive all our results. We note that by virtue of the *inclusion-exclusion principle* [17] and (2.1) one obtains

$$Pr\{D_n > k\} = Pr\{\bigcup_{j=2}^{n} [C_{1j} > k]\} = \frac{1}{n} \sum_{r=2}^{n} (-1)^r \binom{n}{r} \cdot r \cdot Pr\{C_{12} > k, \ldots, C_{1r} > k\}. \tag{2.2}$$

This formula holds for *all* distributions of symbols in the alphabet, and in fact, it can even be extended to dependent keys [18]. Moreover, (2.2) reveals some interesting facts that lead to a better understanding of the depth D_n. In particular, due to the alternating sum in (2.2) we need *exact* formula on the joint distribution $Pr\{C_{12} > k, \ldots, C_{1r} > k\}$ for the asymptotics of the depth D_n. Any approximation of this probability may lead to major difficulties in handling the sum in (2.2).

To illustrate the power and beauty of (2.2), we briefly present analysis of the depth for the independent model, that is, under the following assumption (binary alphabet is assumed only for simplicity of presentation):

A1′ Symbols of the alphabet $\mathcal{A} = \{\omega_1, \omega_2\}$ are generated independently with probability p and $q = 1 - p$ respectively.

Assumption A1′ implies immediately (cf. [16])

$$Pr\{C_{12} > k, \ldots, C_{1r} > k\} = (p^r + q^r)^{k+1} \tag{2.3}$$

and then the generating function $G_n(z) = \sum_{k=0}^{\infty} Pr\{D_k = k\} z^k$ of the depth becomes

$$G_n(z) = 1 - \frac{1-z}{n} \sum_{r=2}^{n} (-1)^r \binom{n}{r} \frac{r(p^r + q^r)}{1 - z(p^r + q^r)} \tag{2.4}$$

for $|z| < 1$. We note that the alternating sum is still in (2.4). In particular, the average depth ED_n follows from (2.4), and after simple algebra one obtains

$$ED_n = \frac{1}{n} \sum_{r=2}^{n} (-1)^r \binom{n}{r} \frac{r(p^r + q^r)}{1 - p^r - q^r}. \tag{2.5}$$

The alternating sums (2.4) and (2.5) are easy to analyze through the Mellin transform as shown in [19] (see also [8, 10-12]), and one immediately obtains asymptotics for ED_n, the variance $var\ D_n$, and limiting distribution of D_n (cf. [9,10,12,13]). We note, however, that our derivations are simple and appealing, and in addition they work without any significant changes for more general models as shown below.

Now we are ready to present our main results for the Markovian model, that is, under assumptions A1 to A3. We note that the above derivation uses the independence assumption A1′ only in obtaining the joint distribution (2.3). In the case of the Markovian model, one easily shows that (2.3) should be replaced by (cf. [20])

$$Pr\{C_{12} > k, \ldots, C_{1r}\} = \sum_{j_1, \ldots, j_{k+1}} [\pi_{j_1}, p_{j_1 j_2}, \ldots, p_{j_k j_{k+1}}]^r = <\pi_{[r]}^k | P_{[r]}^k | \psi_{[r]}> \tag{2.6}$$

where $\pi_{[r]} = (\pi_1^r, \ldots, \pi_V^r)$ and $\pi = (\pi_1, \ldots, \pi_V)$ is the steady-state probability vector of the analyzed Markov chain with transition matrix $P = [p_{ij}]_{i,j=1}^V$. In (2.6) $P_{[r]}$ denotes the r-th Schur product (i.e., elementwise product) of the stochastic matrix P. Furthermore $\psi = (1,1, \ldots, 1)$ is the right eigenvector of P and $\psi_{[r]} = (1^r, 1^r, \ldots, 1^r)$. Finally, $<x,y>$ is the inner product of vector x and y, and $<\pi_{[r]} | P_{[r]}^k | \psi_{[r]}> = <\pi_{[r]}, P_{[r]}^k \psi_{[r]}> = <\pi_{[r]} P_{[r]}^k, \psi_{[r]}>$. Then, (2.6) and our general formula (2.2) immediately implies the generating function $G_n(z)$ of the depth D_n, namely

$$G_n(z) = 1 - \frac{1-z}{n} \sum_{r=2}^{n} (-1)^r \binom{n}{r} \cdot r \cdot <\pi_{[r]} \mid (I - zP_{[r]})^{-1} \mid \psi_{[r]}> \tag{2.7}$$

where I is the identity matrix. We note that (2.7) resembles the generating function (2.4) for the independent model, hence one may expect similar results, and this is confirmed in the proposition below containing our main findings.

We summarize our main results in the following proposition. We delay all proofs to Section 3. We need some notation to present the results in a compact form. Let $P_{[1-z]}$ denote the $(1-z)$-th Schur product of the matrix P, that is, $P_{[1-z]} = \{p_{ij}^{1-z}\}_{i,j=1}^{V}$. By $\pi(z)$ and $\psi(z)$ we denote the left and the right principal eigenvectors of $P_{[1-z]}$, and $\lambda_{[1-z]}$ represents the principal eigenvalue of $P_{[1-z]}$. Finally, $\dot{\pi}(a)$, $\dot{\psi}(a)$, $\dot{P}_{[1-a]}$, $\dot{\lambda}_{[1-a]}$ denote the first derivatives at point a of $\pi(z)$, $\psi(z)$, $P_{[1-z]}$ and $\lambda_{[1-z]}$ respectively.

PROPOSITION. Assume A1–A3 hold, and the Markov chain is stationary.

(i) The average depth ED_n becomes for large n

$$ED_n = \frac{1}{h_1}\{\log n + \gamma + \frac{h_2}{2h_1} - H + P_1(n)\} + O(1/n) \tag{2.8}$$

where $\gamma = 0.577...$ is the Euler constant, $P_1(n)$ is a fluctuating function with a small amplitude (see Section 3), "log" denotes the natural logarithm, and

$$h_1 = -\dot{\lambda}_{[1]} = -<\pi \mid \dot{P}_{[1]} \mid \psi> = -\sum_{i=1}^{V}\sum_{j=1}^{V} \pi_i p_{ij} \log p_{ij} \tag{2.9a}$$

$$h_2 = \ddot{\lambda}_{[1]} = <\pi \mid \ddot{P}_{[1]} \mid \psi \mid > - < \dot{\pi}(1) \mid \dot{P}_{[1]} \mid \psi> - < \pi \mid \dot{P}_{[1]} \mid \dot{\psi}(1)> \tag{2.9b}$$

and $H = -<\dot{\pi}_{[1]}, \psi> = -\sum_{i=1}^{V} \pi_i \log \pi_i$. In the above, $\pi = \pi(1)$ and $\psi = \psi(1) = (1,1,\ldots,1)$ are the left and right eigenvectors of P, that is, $\pi = (\pi_1, \ldots, \pi_V)$ is the stationary distribution.

(ii) The variance $var\ D_n$ of the depth asymptotically becomes

$$var\ D_n = \frac{h_2 - h_1^2}{h_1^3} \log n + c + P_2(n) + O(1/n) \tag{2.10}$$

where c is a constant and $P_2(n)$ is a fluctuating function with small amplitude. For *symmetric independent model* the variance reduces to the constant ($h_2 = h_1^2$) and [11]

$$var\, D_n = \frac{\pi^2}{6 \log^2 V} + \frac{1}{12} + P_2(n) + O(1/n). \tag{2.11}$$

(iii) For the *symmetric independent* model, the limiting distribution of the depth D_n can be expressed as

$$\lim_{n \to \infty} Pr\{D_n < \log_V(xn)\} = e^{-x^{-1}} \tag{2.12}$$

uniformly in $x > 0$. Otherwise, $(D_n - ED_n)/\sqrt{\mathrm{var}\, D_n}$ is asymptotically *normal* with mean zero and variance one, and all moments of D_n tends to the appropriate moments of the limiting normal distribution. ■

Remarks

(i) *Independent model.* The above results trivially reduce to independent models. Indeed, for independent models, one assumes that all rows of the matrix P are the same and equal to $[p_1, p_2, \ldots, p_V]$. Then, the stationary vector π becomes $[p_1, p_2, \ldots, p_V]$ and the entropy $h_1 = -\sum_{i=1}^{V} p_i \log p_i$. It can also be proved that the last two terms in h_2 sum to zero, so $h_2 = \sum_{i=1}^{V} p_i \log^2 p_i$.

(ii) *Higher order Markov chains.* Our analysis also trivially extends to higher order Markov chains. Let us, for simplicity, consider second order Markov chain, that is, the $(\ell + 1)$-st symbol depends on the ℓ-th and the $(\ell - 1)$-st symbols, so the transition probabilities are $p_{ik,j} = Pr\{x_{\ell+1} = \omega_j \,|\, x_\ell = \omega_i,\, x_{\ell-1} = \omega_k\}$. In fact, nothing really new is involved because the sequence of pairs $(x_\ell, x_{\ell+1})$ is an ordinary Markov chain of the first order. Thus, denoting by P the transition matrix of the Markov chain $(x_\ell, x_{\ell+1})$ one immediately obtains our proposition with an appropriate interpretation of h_1 and h_2.

(iii) *Convergence in probability*. Our result can be used to prove convergence of D_n to ED_n *in probability*. Indeed, by Chebyshev's inequality

$$Pr\{\,|D_n/ED_n - 1| \geq \varepsilon\} \leq \frac{\mathrm{var}\, D_n}{\varepsilon^2 (ED_n)^2} = O(1/\log n)$$

so $D_n/ED_n \to 1$ *in probability* as $n \to \infty$. In the case of independent symmetric models, we can prove stronger results, namely *almost surely* convergence of D_n/ED_n to one as $n \to \infty$.

(iv) *Higher moments.* In Proposition (iii) we show that $(D_n - ED_n)/\sqrt{varD_n}$ is asymptotically normal for the asymmetric tries. In fact, we shall prove more, namely that all moments of the above tends to the appropriate moments of the *standard normal distribution* $N(0,1)$. Therefore, one proves that

$$E\left(\frac{D_n - ED_n}{\sqrt{varD_n}}\right)^m \to \begin{cases} \dfrac{m!}{2^{m/2}\,(m/2)!} & m \text{ even} \\ 0 & m \text{ odd} \end{cases}$$

as m tends to infinity. In particular, we may also conclude that the m-th moment ED_n^m of D_n grows asymptotically like $1/h_1^m \cdot \log^m n$ where h_1 is the entropy of the alphabet $\mathcal{A}$.

3. ANALYSIS

In this section we prove our Proposition. The plan is as follows. First, we rederive a more convenient form of the generating function (2.7) for the depth in the Markovian model. Then, we deal in details with the average depth analysis, and shortly discuss the analysis of the variance $var\, D_n$. Finally, we elaborate on the limiting distribution of the depth which is discussed in a fairly detailed way. In this section additionally we assume that the Markov chain is irreducible and aperiodic, and for simplicity of the analysis we adopt another postulate, namely that all eigenvalues are simple† .

† Note that irreducibility of the matrix P implies that the largest eigenvalue is simple. Since only such an eigenvalue really contributes to the asymptotics, we in fact do not limit our analysis.

The generating function $G_n(z)$ of the depth D_n in the Markovian model A1–A3 has been already discussed in Section 2 (cf. (2.6)–(2.7)). It turns out that (2.6) and consequently (2.7), are not in the most convenient form. Let us look again at the joint probability $Pr\{C_{12} > k, \ldots, C_{1r} > k\}$ which is proved in Section 2 to be equal $<\pi_{[r]} | P_{[r]}^k | \psi_{[r]}>$. We recall that for any vector $\mathbf{x}$ by $\mathbf{x}_{[r]}$ we denote the following vector $(x_1^r, x_2^r, \ldots, x_V^r)$. As before $P_{[r]}$ is the r-th Schur product of P and $P_{[r]}^k$ is the k-th power of the matrix $P_{[r]}$. We note that the following spectral representation of $P_{[r]}$ is well known [20]

$$P_{[r]}\mathbf{x} = \lambda_{[r]}^k < \pi, \mathbf{x} > \psi + \sum_{i=2}^{V} g_i(k) \mu_i^k < \pi_i, \mathbf{x} > \psi_i \tag{3.1}$$

where $\mathbf{x}$ is a vector, $\lambda_{[r]}$ is the largest eigenvalue of $P_{[r]}$, π (resp. ψ) is the corresponding left (resp. right) principal eigenvector normalized such that $<\pi, \psi> = 1$. The μ_i, $i = 2, \ldots, V$ denote the remaining eigenvalues and $g_i(k)$ is a polynomial in k. When $P_{[r]}$ involves only simple eigenvalues (no elementary divisors), as we shall assume hereafter throughout the paper, then $g_i(k) \equiv 1$ and π_i (resp. ψ_i) are the associated eigenvectors for μ_i.

Using the spectral representation (3.1) with $g_i(k) \equiv 1$ as mentioned above, we can represent the joint distribution of r alignments as below (see (2.6))

$$\begin{aligned} Pr\{C_{12} > k, \ldots, C_{1r} > k\} &= <\pi_{[r]} | P_{[r]} | \psi_{[r]}> = \\ &= \lambda_{[r]}^k < \pi, \psi_{[r]} > \cdot <\pi_{[r]}, \psi> + \sum_{i=2}^{V} \mu_i^k < \pi_i, \psi_{[r]} > \cdot < \pi_{[r]}, \psi_i > \end{aligned} \tag{3.2}$$

This is the form of the joint probability $Pr\{C_{12} > k, \ldots, C_{1r} > r\}$ that is used in our general formula (2.2) to rederive the generating function $G_n(z)$ of the depth. After some algebra one obtains

$$G_n(z) = 1 - \frac{1-z}{n} \sum_{r=1}^{n} (-1)^r \binom{n}{r} r \left\{ \frac{<\pi, \psi_{[r]}><\pi_{[r]}, \psi>}{1 - z\lambda_{[r]}} + \sum_{i=2}^{V} \frac{<\pi_i, \psi_{[r]}><\pi_{[r]}, \psi_i>}{1 - z\mu_i(r)} \right\}. \tag{3.3}$$

We show soon, the term with eigenvalues μ_i does not contribute at all to the asymptotics.

At first, we deal with the average depth ED_n. Either from (3.3) (e.g., $ED_n = G_n'(1)$) or directly from (2.2) we show that

$$ED_n = \frac{1}{n} \sum_{r=2}^{k} (-1)^r \binom{n}{r} \cdot r \cdot \left\{ \frac{<\pi, \psi_{[r]}><\pi_{[r]}, \psi>}{1-\lambda_{[r]}} + \sum_{i=2}^{V} \frac{<\pi_i, \psi_{[r]}><\pi_{[r]}, \psi_i>}{1-\mu_i(r)} \right\}. \tag{3.4}$$

The asymptotics of the above are easy to establish through the Mellin-like approach suggested by Szpankowski in [19]. It is proved there that an alternating sum of the general form as follows $\sum_{r=2}^{n} (-1)^r \binom{n}{r} \binom{r}{t} f(r)$, where $f(r)$ is any sequence that has analytical continuation to a complex function $f(z)$, can be represented as the following complex integral

$$\sum_{r=2}^{n} (-1)^r \binom{n}{r} \binom{r}{t} f(r) = \frac{1}{2\pi i} \frac{(-1)^t}{t!} \int_{-1/2 - i\infty}^{-1/2 + i\infty} \Gamma(z) f(t-z) n^{t-z} dz + e_n \tag{3.5}$$

where the error function e_n is of at least the order of magnitude smaller than the leading term. More precisely,

$$e_n = O(1/n) \cdot \frac{1}{2\pi i} \int_{-1/2 - i\infty}^{-1/2 + i\infty} z\Gamma(z) f(t-z) n^{t-z} dz \tag{3.5a}$$

where $\Gamma(z)$ is the gamma function [21]. In our case, the formula (3.4) on the average depth ED_n admits the alternating sum form, so by (3.5) we obtain

$$ED_n = -\frac{1}{2\pi i} \int_{-1/2 - i\infty}^{-1/2 + i\infty} \Gamma(z) n^{-z} \left\{ \frac{<\pi(z), \psi_{[1-z]}><\pi_{[1-z]}, \psi(z)>}{1-\lambda_{[1-z]}} + \sum_{i=2}^{V} \frac{<\pi_i(z), \psi_{[1-z]}><\pi_{[1-z]}, \psi_i(z)>}{1-\mu_i(z)} \right\} + e_n/n \tag{3.6}$$

where $\lambda_{[1-z]}$ is the principal eigenvalue of the matrix $P_{[1-z]}$ and $\pi(z)$, $\psi(z)$ are the left and the right principal eigenvectors of $P_{[1-z]}$. From the analyses below, it will be clear that the error e_n becomes $O(1)$.

The evaluation of the integral (3.6) is simple and we appeal to the celebrated Cauchy's residue theorem [21]. We first show that the second term in (3.6) contributes nothing to the asymptotics. Indeed, for this term, right to the line of integration $(-1/2 - i\infty ; -1/2 + i\infty)$ there is only one singularity at $z = 0$ coming from the gamma function. Hence

$$-\frac{1}{2\pi i}\int_{-1/2-i\infty}^{-1/2+i\infty}\Gamma(z)n^{-z}\sum_{i=z}^{V}\frac{<\pi_i(z),\psi_{[1-z]}><\pi_{[1-z]},\psi_i(z)>}{1-\mu_i(z)} = \sum_{i=z}^{V}\frac{<\pi_i(0),\psi><\pi,\psi_i(0)>}{1-\mu_i(0)} \qquad (3.7)$$

But, $<\pi_i(0), \psi> = 0$. Indeed $\pi_i(0) = \pi_i$ is the i-th eigenvector of P, and ψ is the right principal eigenvector of P. Then, $<\pi_i|P|\psi> = \mu_i <\pi_i,\psi> = \lambda <\pi_i,\psi>$, and since $\mu_i \neq \lambda$ we have $<\pi_i,\psi> = 0$. This implies that the sum in (3.7) becomes zero, so the second term of (3.6) does not contribute to the asymptotics.

Let us now concentrate on the first term of (3.6). Here the situation is different since $z = 0$ is also the pole of the denominator, since $\lambda_{[1]} = 1$. Therefore, we have double pole at $z = 0$, and in addition, we have single poles z_k, $k = \pm 1, \pm 2, \ldots,$ at zeros of the denominator, that is, z_k are roots of the following equation

$$\lambda_{[1-z_k]} = 1. \qquad (3.8)$$

We need Taylor's expansions of the function involved in (3.6) to compute the residues. Naturally [21], $\Gamma(z) = z^{-1} - \gamma + O(z)$, and $n^{-z} = 1 - z\log n + O(z^2)$. Moreover, it is easy to notice that

$$<\pi(z),\psi_{[1-z]}><\pi_{[1-z]},\psi(z)> = 1 + H\cdot z + O(z^2) \qquad (3.9a)$$

where $H = -<\dot{\pi}_{[1]}, \psi(1)> = -\sum_{i=1}^{V}\pi_i\log\pi_i$. The denominator $1-\lambda_{[1-z]}$ we handle as follows. Note first that $\lambda_{[1-z]} = <\pi(z)|P_{[1-z]}|\psi(z)>$ and we use the fact that $<\pi(z), \psi(z)> = 1$. So, elementary algebra reveals that

$$\frac{1}{1-\lambda_{[1-z]}} = \frac{1}{h_1 z} + \frac{h_2}{2h_1^2} + O(z) \qquad (3.9b)$$

where h_1 and h_2 are defined in (2.9), that is, $h_1 = -\dot{\lambda}_{[1]}$ and $h_2 = \ddot{\lambda}_{[1]}$. The poles z_k can be handled in a similar manner, and they contributes to a fluctuating function $P_2(x)$. For more details see below. This completes the proof of Proposition (i) formula (2.8).

The variance $var\, D_n$ can be analyzed in a similar manner, and details are left to the reader. We note that from the generating function (3.3) we obtain

$$ED_n(D_n - 1) = \frac{2}{n} \sum_{r=2}^{n} (-1)^r \binom{n}{r} r \frac{<\pi, \psi_{[r]}><\pi_{[r]}, \psi> \cdot \lambda_{[r]}}{(1-\lambda_{[r]})^2} + O(1) \qquad (3.10)$$

and, as above, the asymptotics of (3.10) can be studied throughout the following Mellin-like integral

$$ED_n(D_n - 1) = -\frac{1}{\pi i} \int_{-1/2 - i\infty}^{-1/2 + i\infty} \frac{\Gamma(z) n^{-z} <\pi(z), \psi_{[1-z]}><\pi_{[1-z]}, \psi(z)> \cdot \lambda_{[1-z]}}{(1-\lambda_{[1-z]})^2} dz + O(1/n)$$

Then, $var\, D_n = ED_n(D_n - 1) + ED_n - (ED_n)^2$, and after some algebra one obtains Proposition (ii) formula (2.10).

Finally, we deal with the limiting distribution, and we shall additionally assume that the independent symmetric case is excluded. We use Goncharov's theorem [2] which states that a sequence X_n of random variables with mean μ_n and variance σ_n approaches a normal distribution if

$$\lim_{n\to\infty} e^{-\tau \cdot \mu_n/\sigma_n} G_n(e^{\tau/\sigma_n}) = e^{\tau^2/2}$$

for all $\tau = i\upsilon$ and $-\infty < \upsilon < \infty$, where $G_n(z)$ is the generating function of X_n. Since we are also interested in the convergence in moments, hence we assume that τ is complex. Let $t = \tau/\sigma_n$. Note that $t \to 0$ as $n \to \infty$, since $\sigma_n = \sqrt{\text{var}\, D_n} = \sqrt{\alpha \log n} + O(1)$ where $\alpha = (h_2 - h_1^2)/h_1^3 = \dot\lambda_{[1]} - \ddot\lambda_{[1]} / \dot\lambda^3_{[1]}$. Also, $e^{-\tau \cdot \mu_n/\sigma_n} = \exp\left[-1/h_1 \cdot \sqrt{\log n/\alpha}\right] \to 0$ as $n \to \infty$, since $\mu_n = ED_n = 1/h_1 \log n + O(1)$.

To prove (3.11) we consider first $G_n(\upsilon)$ for $\upsilon = e^t$ with $t \to 0$ for $n \to \infty$. In fact, we may consider $1 - G_n(\upsilon)$. Then, by (3.3)

$$1 - G_n(\upsilon) = \frac{1-\upsilon}{n} \sum_{r=2}^{n} (-1)^n \binom{n}{r} r \frac{<\pi, \psi_{[r]}><\pi_{[r]}, \psi>}{1 - \upsilon\lambda_{[r]}} + O(\upsilon^2) \qquad (3.11)$$

But, (3.11) is an alternating sum with respect to n, so we can evaluate the asymptotics of $1 - G_n(\upsilon)$ by the same approach as in the case of ED_n, that is, by applying our Mellin-like formula (3.5). In our case we have

$$1 - G_n(\upsilon) = -\frac{1-\upsilon}{2\pi i} \int_{-1/2-i\infty}^{-1/2+i\infty} \Gamma(s) n^{-s} \frac{<\pi(s), \psi_{[1-s]}><\pi_{[1-s]}, \psi(s)>}{1 - \upsilon\lambda_{[1-s]}} + O(\upsilon^2) \quad (3.12)$$

To evaluate this integral we apply the Cauchy's residue theorem. For this we need singularities of the function under integral. In our case all of them reduce to the roots of the denominator, that is

$$1 - \upsilon\lambda_{[1-s]} = 0 \quad (3.13)$$

We denote all the roots of (3.13) by $s_k(t)$ when $\upsilon = e^t$, and $k = 0, \pm 1, \pm 2, \ldots$ Then, by Cauchy's formula

$$1 - G_n(\upsilon) = R_0(t) g(s_0(t))(1 - e^t) n^{-s_0(t)} + (1 - e^t) \sum_{\substack{k=-\infty \\ k \neq 0}}^{\infty} R_k(t) g(s_k(t)) n^{-s_k(t)} + O(1) \quad (3.14)$$

where $R_k(t)$ are residues of the denominator, i.e., $[1 - \upsilon\lambda_{[1-s]}]^{-1}$, and $g(s) = \Gamma(s) < \pi(s), \psi_{[1-s]}><\pi_{[1-s]}, \psi(s)>$.

Using the same approach as in Jacquet and Régnier [9] we easily prove that

$$s_0(t) = \frac{t}{\dot\lambda_{[1]}} + \left(\frac{\ddot\lambda_{[1]} - \dot\lambda^2_{[1]}}{\dot\lambda^3_{[1]}} \right) \frac{t^2}{2} + O(t^3)$$

Then, with $t = \tau/\sigma_n = \tau/\sqrt{\alpha \log n}$

$$n^{-s_0(t)} = \exp\left\{ -\frac{\tau}{\dot\lambda_{[1]}} \sqrt{\log n/\alpha} + \tau^2/2 + O(t^3) \right\}$$

We also note that $R_0(t) = -1/h_1 + O(t)$, $g(s_0(t)) = -h_1/t + O(1)$ and $1 - e^t = t + O(t^2)$, so $A = R_0(t) g(s_0(t))(1 - e^t) \underset{t \to 0}{\to} 1$. This will give us the desired proof provided we show that the second term of (3.14) involving $s_k(t)$ is small. But, we note that $\sum_{k \neq 0} R_k(t) g(s_k(t)) = O(1)$, since $R_k(t)$ involves the gamma function (see [9] for details). Then, we have the following estimate

[9] (see also above)

$$|1 - G_n(e^t)| \le n^{-\mathrm{Re}\{s_0(t)\}} A + O(tn^{-\mathrm{Re}\{s_k(t)\}})$$

But, in the Appendix, we prove that $\mathrm{Re}\{s_k(t)\} \ge s_0(\mathrm{Re}(t))$, and this implies for some $\beta > 0$

$$|1 - G_n(e^t)| \le n^{-\mathrm{Re}\{s_0(t)\}} \{A + O(tn^{\mathrm{Re}[s_0(t)] - \mathrm{Re}\{s_0(t)\}})\}$$

$$= n^{-\mathrm{Re}[s_0(t)]} \{A + O(tn^{-\beta t^2})\} \underset{t \to 0}{\to} n^{-\mathrm{Re}\{s_0(t)\}},$$

since $s_0(t) = \frac{t}{\dot{\lambda}_{[1]}} + O(t^2)$. This proves that the contribution from the infinite sum in (3.14) is $n^{-s_0(t)} o(1)$, and finally $e^{-\tau \cdot \mu_n / \sigma_n} G_n(e^{\tau/\sigma_n}) = e^{\tau^2/2} \cdot (1 + o(1))$ as desired. We note that τ was assumed complex, so $G_n(e^t)$ converges to an analytical function, and as such it has well defined derivatives. Hence, we have also proved convergences of all moments.

APPENDIX

We prove in this appendix the following theorem

Theorem. Let $s_k(t)$ be the roots of

$$e^t \lambda_{[1-s]} = 1 \tag{A1}$$

where $\lambda_{[1-s]}$ is the principal eigenvalue of $P_{[1-s]}$, where s and t are complex. Then, for every complex t

$$\mathrm{Re}\{s_k(t)\} \ge s_0\{\mathrm{Re}(t)\} \tag{A2}$$

where $s_0(t) = t/\dot{\lambda}_{[1]} + O(t^2)$. ■

In the proof of the theorem we shall use the following lemma.

Lemma. Let A be a positive matrix, and let exist a vector $\mathbf{x} > 0$ and constant β such that

$$A\mathbf{x} \ge \beta \mathbf{x} \tag{A3}$$

for all $\mathbf{x} \ne 0$, where "$\ge$" means elementwise. Then, the principal eigenvalue λ of A satisfies

$$\lambda \ge \beta \tag{A4}$$

Proof. Define π and ψ as the principal left and right eigenvectors of A. Then

$$\lambda <\pi,x> = <\pi|A|x \geq \beta < \pi,x>$$

so $\lambda \geq \beta$, since $<\pi,x> > 0$, as needed. ■

Corollary. Let t be real, and $P_{[-t]}$ is the Schur product of a stochastic matrix $P = \{p_{ij}\}$. Let $\lambda_{[-t]}$ be the principal eigenvalue of $P_{[-t]}$. Then $\lambda_{[-t]}$ is an increasing function of t.

Proof. By definition $P_{[-t]}\psi_{[-t]} = \lambda_{[-t]}\psi_{[-t]}$ where $\psi_{[-t]}$ is the right eigenvector. But, $P_{[-(t+\Delta t)]}$ is increasing componentwise, that is, $P_{[-(t+\Delta t)]} \geq P_{[-t]}$ for $\Delta t > 0$. Hence, $P_{[-t(t+\Delta t)]}\psi_{[-t]} \geq \lambda_{[-t]}\ \psi_{[-t]}$, so by our Lemma $\lambda_{[-(t+\Delta t)]} \geq \lambda_{[-t]}$, as desired. ■

Now we are ready to prove our Theorem.

Proof of Theorem. From (A1) we obtain

$$e^t P_{[1-s]}\psi(s) = \psi(s) \tag{A5}$$

Since $|\psi(s)|$ is positive vector, then (A5) implies

$$e^{\mathrm{Re}(t)} P_{[1-Re(s)]} |\psi(s)| \geq |\psi(s)|.$$

This and our Lemma imply $\lambda_{[Re(s)]} \geq e^{-\mathrm{Re}(t)}$. But $\lambda_{[s_0(Re(t))]} = e^{-\mathrm{Re}(t)}$, so $\lambda_{[Re(s)]} \geq \lambda_{[s_0(Re(t))]}$. Our Corollary shows that $\lambda_{[x]}$ for x real is an increasing function of x, so $\mathrm{Re}(s_k(t)) \geq s_0(Re(t))$, as needed. ■

REFERENCES

[1] Aho, A., Hopcroft, J. and Ullman, J., *Data Structures and Algorithms*, Addison-Wesley (1983).

[2] Knuth, D., *The Art of Computer Programming, Sorting and Searching*, Addison-Wesley (1973).

[3] Fagin, R., Nievergelt, J., Pippenger, N. and Strong, H., Extendible Hashing: A Fast Access Method for Dynamic Files, *ACM TODS*, 4, pp. 315–344 (1979).

[4] Capetanakis, J., Tree Algorithms for Packet Broadcast Channels, *IEEE Trans. on Inf. Theory*, 25, pp. 605–615 (1979).

[5] Tsybakov, B. and Mikhailov, V., Free Synchronous Packet Access in a Broadcast Channel With Feedback, *Prob. Inf. Transmissions*, 16, pp. 305–317 (1980).

[6] Fayolle, G., Flajolet, Ph., Hofri, M. and Jacquet, Ph., Analysis of a Stack Algorithm for Random Multiple-Access Communication, *IEEE Trans. on Inf. Theory*, **IT-31**, 2, pp. 244–254 (1985).

[7] Szpankowski, W., On a Recurrence Equation Arising in the Analysis of Conflict Resolution Algorithms, *Stochastic Models*, 3, pp. 89–114 (1987).

[8] Flajolet, Ph., On the Performance Evaluation of Extendible Hashing and Trie Searching, *Acta Informatica* 20, pp. 345–369 (1983).

[9] Jacquet, Ph. and Régnier, M., Limiting Distributions for Trie Parameters, in *Lecture Notes in Computer Science*, 214, pp. 196–210 (1986).

[10] Régnier, M. and Jacquet, D., New Results on the Size of Tries, *IEEE Trans. Inf. Theory*, 35, pp. 203–205 (1989).

[11] Szpankowski, W., Some Results on *V*-ary Asymmetric tries, *Journal of Algorithms*, 9, pp. 224–244 (1988).

[12] Kirschenhofer, P. and Prodinger, H., Some Further Results on Digital Search Trees, in *Automata Languages and Machines (ICALP'86)* (L. Kott, ed.), pp. 177–185, *Springer Lecture Notes in Computer Science* 226 (1986).

[13] Pittel, B., Paths in a Random Digital Tree: Limiting Distributions, *Adv. Appl. Prob.*, 18, pp. 139–155 (1986).

[14] Pittel, B., Asymptotical Growth of a Class of Random Trees, *The Annals of Probab.*, 13, pp. 414–427 (1985).

[15] Régnier, M., Trie Hashing Analysis, *Proc. 4-th Int. Conf. on Data Engineering*, pp. 377-387, Los Angeles (1988).

[16] Szpankowski, W., On the Height of Digital Trees and Related Problems, Purdue University, CSD TR-816 (1988); revised (1989); submitted to a journal.

[17] Feller, W., *An Introduction to Probability Theory and its Applications*, John Wiley & Sons, New York (1988).

[18] Jacquet, P. and Szpankowski, W., Do Not Compress Random Strings Analysis of Suffix Tree, submitted to a conference.

[19] Szpankowski, W., The Evaluation of an Alternating Sum With Applications to the Analysis of Some Data Structures, *Information Processing Letters*, 28, pp. 1319 (1988).

[20] Karlin, S. and Ost, F., Counts of Long Aligned Word Matches Among Random Letter Sequences, *Adv. Appl. Probab.*, 19, pp. 293–351 (1987).

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