Partial Fillup and Search Time in LC Tries

SVANTE JANSON
Uppsala University

AND

WOJCIECH SZPANKOWSKI
Purdue University

Abstract. Andersson and Nilsson introduced in 1993 a level-compressed trie (for short, LC trie) in which a full subtree of a node is compressed to a single node of degree being the size of the subtree. Recent experimental results indicated a “dramatic improvement” when full subtrees are replaced by “partially filled subtrees.” In this article, we provide a theoretical justification of these experimental results, showing, among others, a rather moderate improvement in search time over the original LC tries. For such an analysis, we assume that n strings are generated independently by a binary memoryless source, with p denoting the probability of emitting a “1” (and q = 1 − p). We first prove that the so-called α-fillup level $F_n(\alpha)$ (i.e., the largest level in a trie with $\alpha$ fraction of nodes present at this level) is concentrated on two values with high probability: either $F_n(\alpha) = k_{\alpha} n + O(1)$ is an integer and $\Phi(x)$ denotes the normal distribution function. This result directly yields the typical depth (search time) $D_n(\alpha)$ in the $\alpha$-LC tries, namely, we show that with high probability $D_n(\alpha) \sim C_2 \log \log n$, where $C_2 = 1/|\log(1 - h/\log(1/\sqrt{pq}))|$, for $p \neq q$ and $h = -p \log p - q \log q$ is the Shannon entropy rate. This should be compared with recently found typical depth in the original LC tries, which is $C_1 \log \log n$, where $C_1 = 1/|\log(1 - h/\log(1/\min\{p, 1 - p\}))|$. In conclusion, we observe that $\alpha$ affects only the lower term of the $\alpha$-fillup level $F_n(\alpha)$, and the search time in $\alpha$-LC tries is of the same order as in the original LC tries.

Categories and Subject Descriptors: F.2.2. [Analysis of Algorithms and Problem Complexity]: Nonnumerical Algorithms and Problems—Computations on discrete structures; G.2.1 [Discrete Mathematics]: Combinatorics—Generating functions, counting problems

General Terms: Algorithms

Additional Key Words and Phrases: Digital trees, level-compressed tries, partial fillup, probabilistic analysis, Poissonization, strings, trees

The work of the second author was supported in part by NSF Grants CCR-0208709, CCF-0513636, and DMS-0503742, NIH Grant R01 GM068995-01, and AFOSR Grant FA8655-04-1-3074.

Authors’ addresses: S. Janson, Department of Mathematics, Uppsala University, PO Box 480, SE-751 06 Uppsala, Sweden, e-mail: svante.janson@math.uu.se; W. Szpankowski (corresponding author), Department of Computer Science, Purdue University, West Lafayette, IN 47907, e-mail: spa@cs.purdue.edu.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or direct commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or permissions@acm.org.

© 2007 ACM 1549-6325/2007/11-ART44 $5.00 DOI 10.1145/1290672.1290681 http://doi.acm.org/10.1145/1290672.1290681

ACM Transactions on Algorithms, Vol. 3, No. 4, Article 44, Publication date: November 2007.
1. Introduction

Tries and suffix trees are the most popular data structures on words [Gusfield 1997]. A trie is a digital tree built over, say, $n$, strings (the reader is referred to Knuth [1997], Mahmoud [1992], and Szpankowski [2001] for an in-depth discussion of digital trees). A string is stored in an external node of a trie and the path length to such a node is the shortest prefix of the string that is not a prefix of any other strings (see Figure 1). Throughout, we assume a binary alphabet. Then each branching node in a trie is a binary node. A special case of a trie structure is a suffix tree [Szpankowski 2001], which is a trie built over suffixes of a single string.

Since 1960, tries have been used in many computer science applications, such as searching and sorting, dynamic hashing, conflict resolution algorithms, leader election algorithms, IP address lookup, coding, polynomial factorization, Lempel-Ziv compression schemes, and molecular biology. For example, in the Internet IP address lookup problem [Nilsson 1996; Srinivasan and Varghese 1998] one needs a fast algorithm that directs an incoming packet with a given IP address to its destination. As a matter of fact, this is the longest matching prefix problem, and standard tries are well suited for it. However, the search time is too large. If there are $n$ IP addresses in the database, the search time is $O(\log n)$, and this is not acceptable.

In order to improve the search time, [Andersson and Nilsson 1993; Nilsson 1996] introduced a novel data structure called the level-compressed trie, or for short, LC trie (see Figure 1). In the LC trie we replace the root with a node of degree equal to the size of the largest full subtree emanating from the root (the depth of such a subtree is called the fillup level). This is further carried on recursively throughout the whole trie (see Figure 1).

Some recent experimental results reported in Iivonen et al. [1999], Nilsson and Karlsson [1999], and Nilsson and Tikkanen [2002] indicated a “dramatic improvement” in search time when full subtrees are replaced by “partially fillup subtrees.” In this article, we provide a theoretical justification of these experimental results by considering $\alpha$-LC tries in which one replaces a subtree with the last level only $\alpha$-filled by a node of degree equal to the size of such a subtree (and we continue recursively). In order to understand theoretically the $\alpha$-LC trie behavior, we study here the so-called $\alpha$-fillup level $F_n(\alpha)$ and the typical depth, or search time $D_n(\alpha)$. The $\alpha$-fillup level is the last level in a trie that is $\alpha$-filled, that is, filled up to a fraction at least $\alpha$ (e.g., in a binary trie, level $k$ is $\alpha$-filled if it contains $\alpha 2^k$ nodes). The typical depth is the length of a path from the root to a randomly selected external node, thus represents the typical search time. In this article we analyze the $\alpha$-fillup level and typical depth in an $\alpha$-LC trie in a probabilistic framework when all strings are generated by a memoryless source, with $P(1) = p$ and $P(0) = q := 1 - p$. Among other results, we prove that the $\alpha$-LC trie shows a rather moderate improvement over the original LC tries. We shall quantify this statement.

Tries were analyzed over the last 30 years for memoryless and Markov sources [Devroye 1992; Jacquet and Szpankowski 1991; Knessl and Szpankowski 2004; Knuth 1997; Mahmoud 1992; Pittel 1985, 1986; Szpankowski 1991, 2001]. Pittel
[1985, 1986] found the typical value of the fillup level $F_n$ (i.e., $\alpha = 1$) in a trie built over $n$ strings generated by mixing sources; for memoryless sources

$$F_n \sim \frac{\log n}{\log(1/p_{\text{min}})} = \frac{\log n}{h_{-\infty}},$$

where $p_{\text{min}} = \min\{p, 1 - p\}$ is the smallest probability of generating a symbol and $h_{-\infty} = \log(1/p_{\text{min}})$ is the Rényi entropy of infinite order [Szpankowski 2001].

We let $\log := \log_2$. In the preceding, we write $F_n \overset{p}{\sim} a_n$ to denote $F_n/a_n \rightarrow 1$ in probability, that is, for any $\varepsilon > 0$ we have $\mathbb{P}((1 - \varepsilon)a_n \leq F_n \leq (1 + \varepsilon)a_n) \rightarrow 1$ as $n \rightarrow \infty$.

This was further extended by Devroye [1992], and Knessl and Szpankowski [2004], who proved, among other results, that the fillup level $F_n$ is concentrated on two points $k_n$ and $k_n + 1$, where $k_n$ is an integer

$$\frac{1}{\log p_{\text{min}}^{-1}} (\log n - \log \log \log n) + O(1)$$

for $p \neq 1/2$. The depth in regular tries was analyzed by many authors who proved that whp (with high probability, i.e., with probability tending to 1 as $n \rightarrow \infty$) the depth is about $(1/h) \log n$ (where $h = -p \log p - (1 - p) \log(1 - p)$ is the Shannon entropy rate of the source) and that it is normally distributed when $p \neq 1/2$ [Pittel 1986; Szpankowski 2001].

The typical depth (search time) of the original LC tries was analyzed by [Andersson and Nilsson 1993] and by Devroye [2001] for unbiased memoryless sources (see also [Reznik 2002, 2005]). This was only recently extended to general
memoryless sources by Devroye and Szpankowski [2005], who proved that for $p \neq 1/2$,

$$D_n \overset{p}{\sim} \log \log n - \frac{\log(1 - h/h_{-\infty})}{h_{-\infty}},$$

(2)

where, we recall, $h_{-\infty} = \log(1/p_{\min})$.

In this article we shall prove some rather surprising results. First of all, for $0 < \alpha < 1$ we show that the $\alpha$-fillup level $F_n(\alpha)$ is whp equal either to $k_0$ or $k_0 + 1$, where

$$k_0 = \log_{\sqrt{pq}} n - \frac{\ln(p/q)}{2\ln^{3/2}(1/\sqrt{pq})} \Phi^{-1}(\alpha) \sqrt{n} + O(1),$$

(3)

where $\Phi(x)$ is the standard normal distribution function. As a consequence, we find that if $p \neq 1/2$, the depth $D_n(\alpha)$ of the $\alpha$-LC is for large $n$ typically about

$$\log \log n - \frac{\sqrt{\ln n}}{\log(1 - h/\log(1/\sqrt{pq}))}.$$ 

The (full) 1-fillup level (i.e., $\alpha = 1$) $F_n$ shown in Eq. (1) should be compared to the $\alpha$-fillup level $F_n(\alpha)$ presented in (3). Observe that the leading term of $F_n(\alpha)$ is not the same as the leading term of $F_n$ when $p \neq 1/2$. Furthermore, $\alpha$ contributes only to the second term asymptotics. When comparing the typical depths $D_n$ and $D_n(\alpha)$ we conclude that both grow like $\log \log n$ with two constants that do not differ by much (see Figure 2). This comparison led us to a statement in the abstract that the improvement of $\alpha$-LC tries over the regular LC tries is rather moderate. We may add that for relatively slowly growing functions such as $\log \log n$ the constants in front of them do matter (even for large values of $n$) and perhaps this led the authors of Ivonen et al. [1999], Nilsson and Karlsson [1999], and Nilsson and Tikkanen [2002] to their statements.
The article is organized as follows. In the next section we present our main results which are proved in the next two sections. We first consider a Poissonized version of the problem for which we establish our findings. Then we show how to dePoissonize our results, completing our proof for the $\alpha$-fillup. In the last section we prove our second main result concerning the depth.

2. Main Results

Consider tries created by inserting $n$ random strings of 0 and 1. We will always assume that the strings are (potentially) infinite and that the bits in the strings are independent random bits, with $P(1) = p$ and thus $P(0) = q := 1 - p$; moreover, we assume that different strings are independent.

We let $X_k := \#\{\text{internal nodes filled at level } k\}$ and $\overline{X}_k := X_k/2^k$, that is, the proportion of nodes filled at level $k$. Note that $X_k$ may both increase and decrease as $k$ grows, while

$$1 \geq \overline{X}_k \geq \overline{X}_{k+1} \geq 0.$$  

Recall that the fillup level of the trie is defined as the last full level, namely $\max\{k : \overline{X}_k = 1\}$, while the height is the last level with any nodes at all, namely $\max\{k : \overline{X}_k > 0\}$. Similarly, if $0 < \alpha \leq 1$, the $\alpha$-fillup level $F_n(\alpha)$ is the last level where at least a proportion $\alpha$ of the nodes are filled, that is,

$$F_n(\alpha) = \max\{k : \overline{X}_k \geq \alpha\}.$$  

We will in this article study the $\alpha$-fillup level for a given $\alpha$ with $0 < \alpha < 1$ and a given $p$ with $0 < p < 1$.

We have the following result, where whp means with probability tending to 1 as $n \to \infty$, and $\Phi$ denotes the normal distribution function. Theorem 1 is proved in Section 4, after first considering a Poissonized version in Section 3.

**Theorem 1.** Let $\alpha$ and $p$ be fixed with $0 < \alpha < 1$ and $0 < p < 1$, and let $F_n(\alpha)$ be the $\alpha$-fillup level for the trie formed by $n$ random strings as before. Then, for each $n$, there is an integer

$$k_n = \log_{1/\sqrt{pq}} n - \frac{|\ln(p/q)|}{2 \ln^{3/2}(1/\sqrt{pq})} \Phi^{-1}(\alpha) \sqrt{\ln n} + O(1)$$

such that whp $F_n(\alpha) = k_n$ or $k_n + 1$. Moreover, $\mathbb{E} \overline{X}_{k_n} = \alpha + O(1/\sqrt{\ln n})$ for $p \neq 1/2$.

Thus the $\alpha$-fillup level $F_n(\alpha)$ is concentrated on at most two values; as in many similar situations [Devroye 1992; Knesl and Szpankowski 2004; Pittel 1985; Szpankowski 2001], it is easily seen from the proof that in fact, for most $n$, it is concentrated on a single value $k_n$, but there are transitional regimes close to the values of $n$ where $k_n$ changes, where $F_n(\alpha)$ takes two values with comparable probabilities.

Note that when $p = 1/2$, the second term on the righthand side disappears, and thus simply $k_n = \log n + O(1)$; in particular, two different values of $\alpha \in (0, 1)$ have their corresponding $k_n$ differing by $O(1)$ only. When $p \neq 1/2$, changing $\alpha$ means shifting $k_n$ by $\Theta(\log^{1/2} n)$. By Theorem 1, whp $F_n(\alpha)$ is shifted by the same amounts.
To the first order, we thus have the following simple result.

COROLLARY 2. For any fixed $\alpha$ and $p$ with $0 < \alpha < 1$ and $0 < p < 1$, 
\[ F_n(\alpha) = \log \frac{1}{\sqrt{pq}} n + O_p(\sqrt{\ln n}). \]

In particular, 
\[ F_n(\alpha)/\log_{1/\sqrt{pq}} n \xrightarrow{p} 1 \quad \text{as} \quad n \to \infty. \]

Surprisingly enough, the leading terms of the fillup level for $\alpha = 1$ and $\alpha < 1$ are quantitatively different for $p \neq 1/2$. It is well known, as explained in the Introduction, that the regular fillup level $F_n$ is concentrated on two points around $\log n / \log(1/p_{\min})$, while the partial fillup level $F_n(\alpha)$ concentrates around $\log n / \log(1/\sqrt{pq})$. Secondly, the leading term of $F_n(\alpha)$ does not depend on $\alpha$ and the second term is proportional to $\sqrt{\log n}$, whereas for the regular fillup level $F_n$, the second term is of order $\log \log \log n$.

A formal proof of our main result is presented in the next two sections where we first consider a Poissonized version of the problem (see Section 3) followed by dePoisonization (Section 4). However, before proceeding we present a heuristic argument leading to the first term of $F_n(\alpha)$. First, observe that among all $2^k$ binary strings of length $k$, most of them have about $k/2$ zeroes and $k/2$ ones. Thus, the probability that one of the input strings begins with a particular string of length $k$ is usually about $(pq)^{k/2} = (\sqrt{pq})^k$, and the average number of such strings is about $n(\sqrt{pq})^k$. To find the first term of $F_n(\alpha)$ we just set $n(\sqrt{pq})^k = \Theta(1)$, leading to 
\[ F_n(\alpha) \sim \log \frac{1}{\sqrt{pq}} n. \]

In passing, we observe that when analyzing 1-fillup, we have to consider the worst case with all 0’s or all 1’s, which occurs with probability $p_{\min}^k$. Theorem 1 yields several consequences for the behavior of $\alpha$-LC tries. In particular, it implies the typical behavior of the depth, that is, the search time. Next we formulate our second main result concerning the depth for $\alpha$-LC tries, delaying the proof to Section 5; see Eq. (2) and Devroye and Szpankowski [2005] and Reznik [2005] for LC tries.

THEOREM 3. For any fixed $0 < \alpha < 1$ and $p \neq 1/2$ we have 
\[ D_n(\alpha) \overset{p}{\sim} \frac{\log \log n}{-\log \left( 1 - \frac{h}{\log(1/\sqrt{pq})} \right)} \] 
as $n \to \infty$, where $h = -p \log p - (1 - p) \log(1 - p)$ is the entropy rate of the source.

As a direct consequence of Theorem 3, we can numerically quantify the experimental results recently reported in Nilsson and Karlsson [1999] where a “dramatic improvement” in the search time of $\alpha$-LC tries over the regular LC tries was observed. In a regular LC trie the search time is $O(\log \log n)$ with the constant in front of $\log \log n$ being $C_1 = 1/\log(1 - h/\log(1/p_{\min}))^{-1}$ [Devroye and Szpankowski 2005]. For $\alpha$-LC tries this constant decreases to $C_2 = 1/\log(1 - h/\log(1/\sqrt{pq}))^{-1}$ (see Figure 2). While it is hardly a “dramatic improvement,” the fact that we deal with a slowly growing leading term $\log \log n$ may indeed lead to experimentally observed significant changes in the search time.
3. Poissonization

In this section we consider a Poissonized version of the problem, where there are Po(λ) strings inserted in the trie. We let \( \tilde{F}_\lambda(\alpha) \) denote the \( \alpha \)-fillup level of this trie.

**Theorem 4.** Let \( \alpha \) and \( p \) be fixed with \( 0 < \alpha < 1 \) and \( 0 < p < 1 \), and let \( \tilde{F}_\lambda(\alpha) \) be the \( \alpha \)-fillup level for the trie formed by Po(λ) random strings as earlier. Then, for each \( \lambda > 0 \) there is an integer

\[
k_\lambda = \log_2 \lambda - \frac{\left| \ln(p/q) \right|}{2 \ln^{3/2}(1/\sqrt{pq})} \Phi^{-1}(\alpha) \sqrt{\ln \lambda} + O(1)
\]

such that whp (as \( \lambda \to \infty \)) \( \tilde{F}_\lambda(\alpha) = k_\lambda \) or \( k_\lambda + 1 \).

We shall prove Theorem 4 through a series of lemmas. Observe first that a node at level \( k \) can be labeled by a binary string of length \( k \), and that the node is filled if and only if at least two of the inserted strings begin with this label. For \( r \in \{0, 1\}^k \), let \( N_1(r) \) be the number of ones in \( r \), and let \( P(r) = p^{N_1(r)} q^{k-N_1(r)} \) be the probability that a random string begins with \( r \). Then, in the Poissonized version, the number of inserted strings beginning with \( r \in \{0, 1\}^k \) has a Poisson distribution Po(\( \lambda P(r) \)), and these numbers are independent for different strings \( r \) of the same length. The independence is a consequence of the Poisson assumption, in particular the fact that splitting a Poisson process leads to independent Poisson processes [Feller 1971].

Thus

\[
X_k = \sum_{r \in \{0, 1\}^k} I_r,
\]

where \( I_r \) are independent indicators with

\[
\mathbb{P}(I_r = 1) = \mathbb{P}(\text{Po}(\lambda P(r)) \geq 2) = 1 - (1 + \lambda P(r)) e^{-\lambda P(r)}.
\]

Hence

\[
\text{Var}(X_k) = \sum_{r \in \{0, 1\}^k} P(I_r = 1)(1 - P(I_r = 1)) < 2^k
\]

so \( \text{Var}(X_k) < 2^k \) and, by Chebyshev’s inequality,

\[
\mathbb{P}(|X_k - \mathbb{E}[X_k]| > 2^{-k/3}) \to 0.
\]

Consequently, \( X_k \) is sharply concentrated, and it is enough to study its expectation. (It is straightforward to calculate \( \text{Var}(X_k) \) more precisely, and to obtain a normal limit theorem for \( X_k \), but we do not need this.)

Assume first that \( p > 1/2 \).

**Lemma 1.** If \( p > 1/2 \) and

\[
k = \log_2 \lambda - \frac{\left| \ln(p/q) \right|}{2 \ln^{3/2}(1/\sqrt{pq})} \Phi^{-1}(\alpha) \sqrt{\ln \lambda} + O(1),
\]

then \( \mathbb{E}[X_k] = \alpha + O(k^{-1/2}) \).
PROOF. Let \( \rho = p/q > 1 \) and define \( \gamma \) by \( \lambda p^\gamma q^{1-\gamma} = 1 \), namely,
\[ \rho^\gamma = \left( \frac{p}{q} \right)^\gamma = \frac{1}{\lambda} q^{-k}, \]
which leads to
\[ \gamma = \frac{k \ln(1/q) - \ln \lambda}{\ln(p/q)} . \] (10)

Let \( \mu_j = \lambda p^j q^{k-j} = \rho^{j-\gamma} \). Thus, \( \mu_j \) is the average number of strings beginning with a given string with \( j \) ones and \( k-j \) zeros. By Eqs. (6) and (7),
\[ \mathbb{E} X_k = 2^{-k} \sum_{j=0}^{k} \binom{k}{j} \mathbb{P}(\text{Po}(\mu_j) \geq 2). \] (11)

If \( j < \gamma \), then \( \mu_j < 1 \) and
\[ \mathbb{P}(\text{Po}(\mu_j) \geq 2) < \mu^2_j < \mu_j. \]

If \( j \geq \gamma \), then \( \mu_j \geq 1 \) and
\[ 1 - \mathbb{P}(\text{Po}(\mu_j) \geq 2) = (1 + \mu_j) e^{-\mu_j} \leq 2 \mu_j e^{-\mu_j} < 4 \mu_j^{-1} . \]

Hence (11) yields, using \( \binom{k}{j} \leq \binom{k}{\lfloor k/2 \rfloor} = O(2^k k^{-1/2}) \),
\[ \mathbb{E} X_k = 2^{-k} \sum_{j<\gamma} \binom{k}{j} \mu_j + 2^{-k} \sum_{j\geq\gamma} \binom{k}{j} (1 - O(\mu_j^{-1})) \]
\[ = 2^{-k} \sum_{j<\gamma} \binom{k}{j} + 2^{-k} \sum_{j=0}^{k} \binom{k}{j} O(\rho^{-|j-\gamma|}) \]
\[ = \mathbb{P}(\text{Bi}(k, 1/2) \geq \gamma) + O(k^{-1/2}). \] (12)

By the Berry–Esseen theorem [Feller 1971, Theorem XVI.5.1],
\[ \mathbb{P}(\text{Bi}(k, 1/2) \geq \gamma) = 1 - \Phi \left( \frac{\gamma - k/2}{\sqrt{k/4}} \right) + O(k^{-1/2}). \] (13)

By Eq. (10) and the assumption (9),
\[ \gamma - \frac{k}{2} = \frac{1}{\ln(p/q)} \left( k \ln \frac{1}{q} - \ln \lambda - \frac{k}{2} \ln \frac{p}{q} \right) \]
\[ = \frac{1}{\ln(p/q)} \left( k \ln \frac{1}{\sqrt{pq}} - \ln \lambda \right) \]
\[ = \frac{\ln(1/\sqrt{pq})}{\ln(p/q)} \left( k - \log_{1/\sqrt{pq}} \lambda \right) \]
\[ = -\frac{1}{2} \left( \ln(1/\sqrt{pq}) \right)^{-1/2} \Phi^{-1}(\alpha) \sqrt{\ln \lambda} + O(1) \]
\[ = -\frac{1}{2} \Phi^{-1}(\alpha) k^{1/2} + O(1). \] (14)
This finally implies
\[ 1 - \Phi\left(\frac{\gamma - k/2}{\sqrt{k/4}}\right) = 1 - \Phi(-\Phi^{-1}(\alpha)) + O(k^{-1/2}) = \alpha + O(k^{-1/2}), \]
and the lemma follows by (12) and (13).

LEMMA 2. Fix \( p > 1/2 \). For every \( A > 0 \), there exists \( c > 0 \) such that if 
\[ |k - \log_{1/\sqrt{pq}} \lambda| \leq Ak^{1/2}, \]
then \( E \bar{X}_k - E \bar{X}_{k+1} > ck^{-1/2} \).

PROOF. A string \( r \in \{0,1\}^k \) has two extensions \( r0 \) and \( r1 \) in \( \{0,1\}^{k+1} \). Clearly, 
\( l_{r0}, l_{r1} \leq l_r \), and if exactly two (or three) of the inserted strings begin with \( r \), then 
\( l_{r0} + l_{r1} \leq 2l_r \). Hence 
\[ E(2X_k - X_{k+1}) = \sum_{r \in \{0,1\}^k} E(2l_r - l_{r0} - l_{r1}) \geq \sum_{r \in \{0,1\}^k} \mathbb{P}(\text{Po}(\lambda, P(r)) = 2). \quad (15) \]
Let \( \rho \) and \( \gamma \) be as in the proof of Lemma 1, and let \( j = \lceil \gamma \rceil \). Then \( \mu_j = \rho^{j-\gamma} \in [1, \rho] \) and thus \( \mathbb{P}(\text{Po}(\mu_j) = 2) \geq \frac{1}{2} e^{-\rho} \). Moreover, by (14) and the assumption, 
\[ |j - k/2| \leq \frac{\ln(1/\sqrt{pq})}{\ln(p/q)} Ak^{1/2} + 1 = O(k^{1/2}). \]
Thus, if \( k \) is large enough, we have by the standard normal approximation of the binomial probabilities (which follows easily from Stirling’s formula, as found already by de Moivre [1738]), then
\[ 2^{-k} \binom{k}{j} = \frac{1 + o(1)}{\sqrt{2\pi k/4}} e^{-2(j-k/2)^2/k} \geq c_1 k^{-1/2} \]
for some \( c_1 > 0 \). Hence by (15),
\[ E \bar{X}_k - E \bar{X}_{k+1} = 2^{-k-1} E(2X_k - X_{k+1}) \geq 2^{-k-1} \binom{k}{j} \mathbb{P}(\text{Po}(\mu_j) = 2) \geq \frac{c_1 e^{-\alpha}}{4} k^{-1/2}, \]
as needed.

Now assume that \( p > 1/2 \). Starting with any \( k \) as in Eq. (9), we can by Lemmas 1 and 2 shift \( k \) up or down \( O(1) \) steps and find \( k_\lambda \) as in (5) such that, for a suitable \( c > 0 \), 
\[ E \bar{X}_{k_\lambda} \geq \alpha + \frac{1}{2} c k^{-1/2} > E \bar{X}_{k_\lambda+1} \quad \text{and} \quad E \bar{X}_{k_\lambda+2} \leq E \bar{X}_{k_\lambda+1} - c k_\lambda^{-1/2} < \alpha - \frac{1}{2} c k_\lambda^{-1/2}. \]
It follows by (8) that whp \( \bar{X}_{k_\lambda} \geq \alpha \) and \( \bar{X}_{k_\lambda+2} < \alpha \), and hence \( \tilde{F}_\lambda(\alpha) = k_\lambda \) or \( k_\lambda + 1 \).

This proves Theorem 4 in the case \( p > 1/2 \). The case \( p < 1/2 \) follows by symmetry, interchanging \( p \) and \( q \).

In the remaining case \( p = 1/2 \), all \( P(r) = 2^{-k} \) are equal. Thus by Eqs. (6) and (7),
\[ E \bar{X}_k = \mathbb{P}(\text{Po}(\lambda 2^{-k}) \geq 2). \quad (16) \]
Given \( \alpha \in (0,1) \), there is a \( \mu > 0 \) such that \( \mathbb{P}(\text{Po}(\mu) \geq 2) = \alpha \). We take \( k_\lambda = \lceil \log(\lambda/\mu) - 1/2 \rceil \). Then, \( \lambda 2^{-k_\lambda} \geq 2^{1/2} \mu \) and thus \( E \bar{X}_{k_\lambda} \geq \alpha_+ \) for some \( \alpha_+ > \alpha \). Similarly, \( E \bar{X}_{k_\lambda+2} \leq \alpha_- \) for some \( \alpha_- < \alpha \), and the result follows in this case too.
4. DePoissonization

To complete the proof of Theorem 1 we must dePoissonize the results obtained in Theorem 4, which we do in this section.

**Proof of Theorem 1.** Given an integer $n$, let $k_n$ be as in the proof of Theorem 4 with $\lambda = n$, and let $\lambda_\pm = n \pm n^{2/3}$. Then $\mathbb{P}(\text{Po}(\lambda_-) \leq n) \to 1$ and $\mathbb{P}(\text{Po}(\lambda_+) \geq n) \to 1$ as $n \to \infty$. By monotonicity, we thus have whp $\tilde{F}_{\lambda_-}(\alpha) \leq F_n(\alpha) \leq \tilde{F}_{\lambda_+}(\alpha)$, and by Theorem 4 it remains only to show that we can take $k_{\lambda_-} = k_{\lambda_+} = k_n$.

Let us now use the notation $X_k(\lambda)$ and $\overline{X}_k(\lambda)$, since we are working with several $\lambda$.

**Lemma 3.** Assume $p \neq 1/2$. Then, for every $k$,

$$\frac{d}{d\lambda} \mathbb{E} X_k(\lambda) = O(\lambda^{-1} k^{-1/2}).$$

**Proof.** We have

$$\frac{d}{d\lambda} \mathbb{P}(\text{Po}(\mu) \geq 2) = \frac{d}{d\mu}((1 - (1 + \mu)e^{-\mu}) = \mu e^{-\mu}$$

and thus, by (11) and the argument in (12),

$$\frac{d}{d\lambda} \mathbb{E} X_k(\lambda) = 2^{-k} \sum_{j=0}^{k} \binom{k}{j} \mu_j e^{-\mu_j} \frac{d\mu_j}{d\lambda}$$

$$= \lambda^{-1} 2^{-k} \sum_{j=0}^{k} \binom{k}{j} \mu_j^2 e^{-\mu_j} = O(\lambda^{-1} \sum_{j=0}^{k} 2^{-k} \binom{k}{j} \min(\mu_j, \mu_j^{-1}))$$

$$= O(\lambda^{-1} k^{-1/2}),$$

which completes the proof. □

By Lemma 3, $|\mathbb{E} \overline{X}_k(\lambda_\pm) - \mathbb{E} \overline{X}_k(n)| = O(n^{-1/3} k^{-1/2}) = o(k^{-1/2})$. Hence, by the proof of Theorem 4, for large $n$, $\mathbb{E} \overline{X}_k_n(\lambda_\pm) \geq \alpha + \frac{1}{5} c k_n^{-1/2}$ and $\mathbb{E} \overline{X}_k_{n+2}(\lambda_\pm) < \alpha - \frac{1}{5} c k_n^{-1/2}$, and thus whp $\tilde{F}_{\lambda_\pm}(\alpha_k) = k_n$ or $k_n + 1$. Moreover, the estimate $\mathbb{E} \overline{X}_k_n = \alpha + O(1/\sqrt{\log n})$ follows easily from the similar estimate for the Poisson version in Lemma 1; we omit the details. This completes the proof of Theorem 1 for $p > 1/2$. The case $p < 1/2$ is again the same by symmetry. The proof when $p = 1/2$ is similar, now using Eq. (16).

5. Proof of Theorem 3

First, let us explain heuristically our estimate for $D_n(\alpha)$. By the Asymptotic Equipartition property [Szpankowski 2001] at level $k_n$ there are about $n 2^{-h k_n}$ strings with the same prefix of length $k_n$ as a randomly chosen one, where $h$ is the entropy. In other words, in the corresponding branch of the $\alpha$-LC trie, we have about $n 2^{-h k_n} \approx n^{1-\alpha}$ strings (or external nodes), where for simplicity $\kappa = h / \log(1/\sqrt{pq})$. In the next level, we shall have about $n^{(1-\kappa)^2}$ external nodes, and so on. In particular, at level $D_n(\alpha)$ we have approximately

$$n^{(1-\kappa)^{O(n(\alpha))}}$$
external nodes. Setting this = Θ(1) leads to our estimate (4) of Theorem 3.

We now make this argument rigorous. We construct an α-LC trie from n random strings ξ1, . . . , ξn and look at the depth Dn(α) of a designated one of them. In principle, the designated string should be chosen at random, but by symmetry, we can assume that it is the first string ξ1.

To construct the α-LC trie, we scan the strings ξ1, . . . , ξn in parallel one bit-at-a-time, and build a trie level-by-level. As soon as the last level is filled less than α, we stop; we are now at level Fn(α) + 1, just past the α-fillup level. The trie above this level, that is, up to level Fn(α), is compressed into one node, and we continue recursively with the strings attached to each node at level Fn(α) + 1 in the uncompressed trie, namely, the sets of strings that begin with the same prefixes of length Fn(α) + 1.

To find the depth Dn(α) of the designated string ξ1 in the compressed trie, we may ignore all branches not containing ξ1; thus, we let Yn be the number of the n strings that agree with ξ1 for the first Fn(α) + 1 bits. Note that we have not yet inspected any later bits. Hence, conditioned on Fn(α) and Yn, the remaining parts of these Yn strings are again independently and identically distributed random strings from the same memoryless source, so we may argue by recursion. The depth Dn(α) equals the number of recursions needed to reduce the number of strings to 1.

We begin by analyzing a single step in the recursion. Let, for notational convenience, κ := h/ log(1/ sqrt(pq)). Note that 0 < κ < 1.

**Lemma 4.** Let ε > 0. Then, with probability 1 − O(n−Θ(1)),

\[
1 − \kappa − \varepsilon < \frac{\ln Y_n}{\ln n} < 1 − \kappa + \varepsilon. \tag{17}
\]

We postpone the proof of Lemma 4, and first use it to complete the proof of Theorem 3. We assume in the following that n is large enough when needed, and that 0 < ε < min(κ, 1 − κ)/2.

We iterate, and let Zj be the number of strings remaining after j iterations; this is the number of strings that share the first j levels with ξ1 in the compressed trie. We have Z0 = n and Z1 = Yn. We stop the iteration when there are less than ln n strings remaining: we thus let τ be the smallest integer such that Zτ < ln n. In each iteration before τ, Eq. (17) holds with error probability O((ln n)−Θ(1)) = O((ln ln n)−2).

Hence, for any constant B, we have whp for every j ≤ min(τ, B ln ln n), with \( \kappa_\pm = \kappa \pm \varepsilon \in (0, 1) \),

\[
1 − \kappa_+ < \frac{\ln Z_j}{\ln Z_{j-1}} < 1 − \kappa_-, \]

or equivalently

\[
\ln(1 − \kappa_+) < \ln \ln Z_j − \ln \ln Z_{j-1} < \ln(1 − \kappa_-). \tag{18}
\]

If \( \tau > \tau_+ := \left\lceil \ln \ln n / \ln(1 − \kappa_-)^{-1} \right\rceil \), we find whp from (18) that

\[
\ln \ln Z_{\tau_+} ≤ \ln \ln Z_0 + \tau_+ \ln(1 − \kappa_-) ≤ 0,
\]

so \( Z_{\tau_+} ≤ e < \ln n \), which violates \( \tau > \tau_+ \). Hence \( \tau ≤ \tau_+ \) whp.
On the other hand, if \( \tau < \tau_- := \left( (1 - \varepsilon) \ln \ln n / \ln(1 - \kappa_+) \right)^{-1} \), then whp by (18)

\[
\ln \ln Z_\tau \geq \ln \ln Z_0 + \tau_- \ln(1 - \kappa_+) \geq \varepsilon \ln \ln n,
\]

which contradicts \( \ln \ln Z_\tau < \ln \ln \ln n \).

Consequently, whp \( \tau_- \leq \tau \leq \tau_+ \); in other words, we need
\[
\frac{\ln \ln n}{\ln(1 - \kappa)} (1 + O(\varepsilon))
\]
iterations to reduce the number of strings to less than \( \ln \ln n \).

Iterating this result once, we see that whp at most \( O(\ln \ln n) \) further iterations are needed to reduce the number to less than \( \ln \ln n \). Finally, the remaining depth then whp is \( O(\ln \ln \ln n) \) even without compression. Hence we see that whp

\[
D_n(\alpha) = \frac{\ln \ln n}{-\ln(1 - \kappa)} (1 + O(\varepsilon)) + O(\ln \ln n).
\]

Since \( \varepsilon \) is arbitrary, Theorem 3 follows.

It remains to prove Lemma 4. Let \( W_k \) be the number of the strings \( \xi_1, \ldots, \xi_n \) that are equal to \( \xi_1 \) for at least their first \( k \) bits. The \( Y_n = W_{F_e(\alpha)+1} \), and thus for any \( A > 0 \),

\[
\mathbb{P}(|\log Y_n - (1 - \kappa) \log n| \geq 2\varepsilon \log n) \leq \mathbb{P}(|F_n(\alpha) - \log_{1/\sqrt{pq}} n| \geq A\sqrt{\ln n})
\]

\[
+ \sum_{|k-1-\log_{1/\sqrt{pq}} n|<A\sqrt{\ln n}} \mathbb{P}(|\log W_k - \log n + h \log_{1/\sqrt{pq}} n| \geq 2\varepsilon \log n).
\]

Lemma 4 thus follows from the following two lemmas, using the observation that

\[ 0 < 1 / \log(1/\sqrt{pq}) < 1 / h. \]

The first lemma is a large deviation estimate corresponding to Corollary 2.

**Lemma 5.** For each \( \alpha \in (0, 1) \), there exists a constant \( A \) such that

\[ \mathbb{P}(|F_n(\alpha) - \log_{1/\sqrt{pq}} n| \geq A\sqrt{\ln n}) = O(1/n). \]

**Proof.** We begin with the Poissonized version, with \( \text{Po}(\lambda) \) strings as in Section 3. Let \( k_\pm = k_\pm(\lambda) := \lfloor \log_{1/\sqrt{pq}} \lambda \pm A\sqrt{\ln \lambda} \rfloor \), and let \( \delta \) be fixed with

\[ 0 < \delta < \min(\alpha, 1 - \alpha). \]

Then, by Lemma 1, if \( A \) is large enough, \( \mathbb{E} X_{k_-} > \alpha + \delta \) and \( \mathbb{E} X_{k_+} < \alpha - \delta \) for all large \( \lambda \). By a Chernoff bound, (8) can be sharpened to

\[ \mathbb{P}(|X_k - \mathbb{E} X_k| > \delta) = O(e^{-\Theta(2^\delta)}) \]
and thus

\[ \mathbb{P}(F_\alpha(a) < k_-) \leq \mathbb{P}(X_{k_-} < a) \leq \mathbb{P}(X_{k_-} - \mathbb{E} X_{k_-} < -\delta) = O(e^{-\Theta(2^\delta - 1)}) = O(e^{-\Theta(\lambda^{(\alpha)/2})}) = O(\lambda^{-1}). \]

Similarly, \( \mathbb{P}(F_\alpha(a) > k_+) = O(\lambda^{-1}) \).

To dePoissonize, let \( \lambda_\pm = n \pm n^{2/3} \) as in Section 4 and note that, again by a Chernoff estimate, \( \mathbb{P}(\text{Po}(\lambda_-) \leq n) = O(n^{-1}) \) and \( \mathbb{P}(\text{Po}(\lambda_+) \geq n) = O(n^{-1}) \). Thus, with probability \( 1 - O(1/n) \),

\[ k_-(\lambda_-) \leq F_\alpha(a) \leq F_n(a) \leq F_\alpha(a) \leq k_+(\lambda_+), \]

and the result follows (if we increase \( A \)).
Lemma 6. Let $0 < a < b < 1/h$ and $\varepsilon > 0$. Then, uniformly for all $k$ with $a \log n \leq k \leq b \log n$,
\[
P(|\log W_k - \log n + kh| > \varepsilon \log n) = O(n^{-\Theta(1)}). \tag{19}
\]

Proof. Let $N_1$ be the number of 1’s in the first $k$ bits of $\xi_i$. Given $N_1$, the distribution of $W_k - 1$ is $\text{Bi}(n - 1, p^{N_1/k}q^{k-N_1})$.

Since $p^q q^k = 2^{-h}$, there exists $\delta > 0$ such that if $|N_1/k - p| \leq \delta$, then $2^{-h-\varepsilon} \leq p^{N_1/k}q^{k-N_1/k} \leq 2^{-h+\varepsilon}$, and thus
\[
2^{-hk-ek} \leq p^{N_1/k}q^{k-N_1} \leq 2^{-hk+ek}, \quad \text{when } |N_1/k - p| \leq \delta. \tag{20}
\]

Noting that $hk \leq bh \log n$ and $bh < 1$, we see that, provided $\varepsilon$ is small enough, $n2^{-hk-ek} \geq n^\eta$ for some $\eta > 0$, and then (20) and a Chernoff estimate yields, when $|N_1/k - p| \leq \delta$,
\[
P\left(\frac{1}{2}n2^{-hk-ek} \leq W_k \leq 2n2^{-hk-ek} \mid N_1\right) = 1 - O\left(e^{-\Theta(n^\eta)}\right) = 1 - O(n^{-1}),
\]
and thus
\[
P(\log W_k - \log n + hk > \varepsilon k + 1 \mid N_1) = O(n^{-1}), \quad \text{when } |N_1/k - p| \leq \delta. \tag{21}
\]

Moreover, $N_1 \sim \text{Bi}(k, p)$, so by another Chernoff estimate,
\[
P(|N_1/k - p| > \delta) = O\left(e^{-\Theta(k)}\right) = O(n^{-\Theta(1)}).
\]
The result follows (possibly changing $\varepsilon$) from this and Eq. (21). \qed

References

Andersson, A., and Nilsson, S. 1993. Improved behavior of tries by adaptive branching. Inf. Proc. Lett. 46, 295–300.

Devroye, L. 1992. A note on the probabilistic analysis of Patricia tries. Random Struct. Alg. 3, 203–214.

Devroye, L. 2001. An analysis of random LC tries. Random Struct. Alg. 19, 359–375.

Devroye, L., and Szpankowski, W. 2005. Probabilistic behavior of asymmetric level compressed tries. Random Struct. Alg. 27, 2, 185–200.

De Moivre, A. 1738. The Doctrine of Chances, 2nd ed., H. Woodfall, London.

Feller, W. 1971. An Introduction to Probability Theory and its Applications. Vol. II. 2nd ed., Wiley, New York.

Gusfield, D. 1997. Algorithms on Strings, Trees, and Sequences. Cambridge University Press, New York.

Iivonen, P., Nilsson, S., and Tikkonen, M. 1999. An experimental study of compression methods for functional tries. In Workshop on Algorithmic Aspects of Advanced Programming Languages (WAAAPL).

Jacquet, P., and Szpankowski, W. 1991. Analysis of digital tries with Markovian dependency. IEEE Trans. Inf. Theory 37, 1470–1475.

Jacquet, P., and Szpankowski, W. 1998. Analytical dePoissonization and its applications. Theor. Comput. Sci. 201, 1–62.

Knessl, C., and Szpankowski, W. 2004. On the number of full levels in tries. Random Struct. Alg. 25, 247–276.

Knuth, D. E. 1997. The Art of Computer Programming. Vol. 1: Fundamental Algorithms, 3rd ed. Addison-Wesley, Reading, MA.

Knuth, D. E. 2000. Selected Papers on Analysis of Algorithms, CSLI, Stanford, CA.

Mahmoud, H. 1992. Evolution of Random Search Trees. John Wiley and Sons, New York.

Nilsson, S. 1996. Radix Sorting and Searching. PhD Thesis, Lund University.

Nilsson, S., and Karlsson, G. 1998. Fast address look-up for Internet routers. In Proceedings of the IFIP 4th International Conference on Broadband Communications, 11–22.
NILSSON, S., AND KARLSSON, G. 1999. IP-address lookup using LC-tries. IEEE J. Select. Areas in Commun. 17, 6, 1083–1092.

NILSSON, S., AND TIKKANEN, M. 2002. An experimental study of compression methods for dynamic tries. Algorithmica 33, 1, 19–33.

PITTEL, B. 1985. Asymptotic growth of a class of random trees. Ann. Probab. 18, 414–427.

PITTEL, B. 1986. Paths in a random digital tree: limiting distributions, Adv. Appl. Probab. 18, 139–155.

REZNİK, Y. 2002. Some results on tries with adaptive branching. Theor. Comput. Sci. 289, 1009–1026.

REZNİK, Y. 2005. On the average density and selectivity of nodes in multi-digit tries. In Proceedings of the 7th Workshop on Algorithm Engineering and Experiments and the 2nd Workshop on Analytic Algorithmics and Combinatorics (ALENEX/ANALCO) (Vancouver, British Columbia, Canada). SIAM, 230–239.

SRINIVASAN, V., AND VARGHESE, G. 1998. Fast address lookups using controlled prefix expansions. ACM SIGMETRICS.

SZPANKOWSKI, W. 1991. On the height of digital trees and related problems. Algorithmica 6, 256–277.

SZPANKOWSKI, W. 2001. Average Case Analysis of Algorithms on Sequences. John Wiley, New York.

RECEIVED OCTOBER 2005; REVISED AUGUST 2006; ACCEPTED OCTOBER 2006