RAIRO-Theor. Inf. Appl. 51 (2017) 7–15
DOI: 10.1051/ita/2017002

Available online at:
www.rairo-ita.org

ANALYSIS OF DIGITAL SEARCH TREES INCORPORATED WITH PAGING *

Mehri Javanian[1] and Mohammad Q. Vahidi-asl[2]

Abstract. Ordinary digital search trees (DSTs) stores one word in each of its internal nodes and leaves, but a DST with paging size b allows storing b words in the leaves, which corresponds to pages in auxiliary storage. In this paper, we analyse the average number of nodes, the average node-wise path length and 2-protected nodes in DSTs with paging size b. We utilize recurrence relations, analytical Poissonization and de-Poissonization, the Mellin transform, and complex analysis. We also compare the storage usage in paged DSTs to that in DSTs. For example, for $b = 2, 3, 4, 5, 6$, the approximate average number of nodes in paged DSTs is, respectively, 82%, 67%, 55%, 47%, 41% of the size of DSTs (when $b = 1$). Thus the results are nontrivial and interesting for computer scientists.

Mathematics Subject Classification. 05C05, 60C05.

1. Introduction

The transfer of pages between main memory and an auxiliary store, such as hard disk drive is called paging. Due to many advantages of paging, a number of authors have studied its implementation on search trees.

A paged binary search tree with page capacity b stores all its subtrees of size $\leq b$ (possibly empty) in pages; typically, the pages reside in secondary memory and the elements within a page are not organized as search trees (for further details see [10, 11]).

It is well-known that we can associate to each particular execution of Quicksort, a binary search tree: the root contains the pivot element of the first stage, and the left and right subtrees are recursively built for the elements smaller and larger than the pivot, respectively. Each internal node in the binary search tree corresponds to a recursive call to Quicksort. It can be made a partitioning of a given subfile if and only if the subfile contains $> b$ elements, *i.e.* the corresponding internal node has $> b$ descendants. Martinez *et al.* [10] have been studied the number of partitions made by Quicksort to sort n elements, when the recursion halts on subfiles of size $\leq b$. They also have applied their results in the study of the number of descendants in the context of paged trees instance and have found the expected number of pages in a random binary search tree of size n with page capacity b. The results obtained for patterns in random binary search trees has been applied to paged trees or to Quicksort with halting on short subfiles (size $\leq b$) by Flajolet *et al.* [3]. Suitable adaptations of their technique also lead to a distributional analysis of paging.

Keywords and phrases. Random trees, paged digital search trees, singularity analysis.

* *The first author would like to thank professor Hsien-Kuei Hwang for suggesting this topic of research and his hospitality during her visit to the Institute of Statistical Sciences, Academia Sinica in Taipei, Taiwan.*

[1] Department of Statistics, University of Zanjan, Zanjan, Iran. javanian@znu.ac.ir

[2] Department of Statistics, Shahid Beheshti University, Tehran, Iran. m-vahidi@sbu.ac.ir

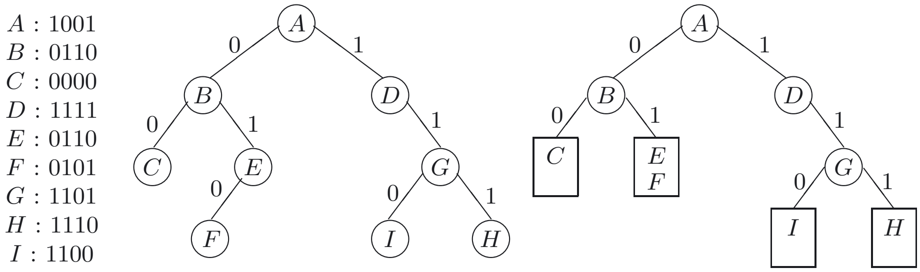

FIGURE 1. A digital search tree (*left*) of size 9, and the corresponding paged digital search tree with page capacity $b = 2$ (*right*): the size of paged tree is 8.

Hoshi and Flajolet [6] have been provided a characterization of the storage needs of a quadtree when used as an index to access large volumes of 2-dimensional data. They have shown that the page occupancy for data in random order approaches 33%. They concentrate on quadtrees that depend on an integer parameter $b \geq 1$ representing a page capacity. Their major results characterize the expected storage occupancy of quadtrees. For data in random order, they establish that the filling ratio of pages is approximately 33%, in the sense that the number of pages necessary to store a file of n points with b the page capacity is about $3n/b$.

The digital tree structure can be extended by letting nodes contain up to b elements, but still retaining the binary branching principle. Vitter and Flajolet [11] have discussed extension of some results including the theorem on the average number of internal nodes in a trie to the case in which each external node in the trie represents a page of secondary storage capable of storing $b \geq 1$ elements. They found the number of pages is about $n/b \log 2$ and the filling ratio of pages is approximately 69%.

In this paper, we study the paging cost associated with the use of Digital Search Trees (DSTs), a structure that is important in Computer Science (see [9]). DSTs are constructed as follows. Given a set of n keys which are infinite 0-1 strings, we place the first key in the root node; those starting with "0" ("1") are directed to the left (right) subtree of the root, and are constructed recursively by the same procedure but with the removal of their first bits when comparisons are made. See Figure 1 for an illustration.

By random DST we mean that whenever a decision has to be made whether to go down to the left or right, a fair coin is tossed, and a direction is chosen with probability $\frac{1}{2}$. Shape parameters in random DSTs, like total path-length, weighted path-length have been analyzed in many papers; see [6] and references therein.

A random paged DST is constructed as a random DST except that it depends on an integer parameter $b \geq 1$ representing a page capacity, small subtrees with size $\leq b$ are represented sequentially into a page instead of being split recursively. Figure 1 illustrates a DST (left) along with its paged version (right) where the pages are shown as boxes and the internal nodes as circles.

2. A LEMMA

We use X_n to denote the node-wise path-length (which is defined as the sum of the distances of all nodes to the roots, regardless of the number of keys in each node) in a random paged DST with page capacity b built from n keys. A node is 2-protected if its distance to any descendant that is a leaf is at least 2. In other words, a 2-protected node is not leaf, neither are its children. Let N_n and Y_n stand for the number of nodes and the number of 2-protected nodes in this kind of trees with n keys.

Obviously, For $n \geq b+1$, X_n and N_n satisfies the following distributional recurrences

$$N_n \stackrel{d}{=} N_{B_{n-1}} + N^*_{n-1-B_{n-1}} + 1,$$
$$X_n \stackrel{d}{=} X_{B_{n-1}} + X^*_{n-1-B_{n-1}} + N_{B_{n-1}} + N^*_{n-1-B_{n-1}},$$

with the initial conditions $N_0 = 0$, $N_j = 1$ and $X_j = 0$ for $1 \leq j \leq b$. Here $X_n \stackrel{d}{=} X^*_n$, $N_n \stackrel{d}{=} N^*_n$, $B_n \stackrel{d}{=}$ Binomial$(n, 1/2)$ and X_n, X^*_n, B_n as well as N_n, N^*_n, B_n are independent. Also for $n \geq b+2$, Y_n satisfies

$$Y_{n+1} \stackrel{d}{=} Y_{B_n} + Y^*_{n-B_n} + 1 - \mathbb{I}_{\{B_n=1\}} - \mathbb{I}_{\{B_n=n-1\}},$$

with the initial conditions $X_j = 0$ for $0 \leq j \leq b+1$, and

$$Y_{b+2} = \begin{cases} 1, & \text{with probability } 2^{-b}; \\ 0, & \text{with probability } 1-2^{-b}; \end{cases}$$

($\stackrel{d}{=}$ and $\mathbb{I}_A$ denote, respectively, equal in distribution and indicator function of event A). Here $Y_n \stackrel{d}{=} Y^*_n$ and Y_n, Y^*_n, B_n are independent.

From the above relations, taking expectation translates into the recurrences

$$\mathbb{E}(N_n) = 2^{-n+2} \sum_{j=0}^{n-1} \binom{n-1}{j} \mathbb{E}(N_j) + 1,$$
$$\mathbb{E}(X_n) = 2^{-n+2} \sum_{j=0}^{n-1} \binom{n-1}{j} \Big(\mathbb{E}(X_j) + \mathbb{E}(N_j)\Big),$$

for $n \geq b+1$; with $\mathbb{E}(N_0) = 0$, $\mathbb{E}(N_j) = 1$ and $\mathbb{E}(X_j) = 0$ for $1 \leq j \leq b$. Also, for $n \geq b+2$

$$\mathbb{E}(Y_{n+1}) = 1 + 2^{-n+1} \sum_{j=0}^{n} \binom{n}{j} \mathbb{E}(Y_j) - n2^{-n+1},$$

with $\mathbb{E}(Y_n) = 0$ for $0 \leq j \leq b+1$, and $\mathbb{E}(Y_{b+2}) = 2^{-b}$.

Lemma 2.1. *The Poisson transforms of* $\mathbb{E}(N_n)$, $\mathbb{E}(X_n)$ *and* $\mathbb{E}(Y_n)$, *namely*

$$\tilde{f}(z) := \mathrm{e}^{-z} \sum_{n\geq 0} \frac{\mathbb{E}(N_n)}{n!} z^n, \qquad \tilde{g}(z) := \mathrm{e}^{-z} \sum_{n\geq 0} \frac{\mathbb{E}(X_n)}{n!} z^n, \qquad \tilde{t}(z) := \mathrm{e}^{-z} \sum_{n\geq 0} \frac{\mathbb{E}(Y_n)}{n!} z^n,$$

fulfills the following functional recurrences

$$\tilde{f}(z) + \tilde{f}'(z) = 2\tilde{f}(z/2) + 1 - \sum_{j=1}^{b-1} \frac{z^j \mathrm{e}^{-z}}{j!} \left(2 - 2^{-j+1}\right), \tag{2.1}$$

$$\tilde{g}(z) + \tilde{g}'(z) = 2\tilde{g}(z/2) + 2\tilde{f}(z/2) - \sum_{j=1}^{b-1} \frac{z^j \mathrm{e}^{-z}}{j!} \left(2 - 2^{-j+1}\right), \tag{2.2}$$

$$\tilde{t}(z) + \tilde{t}'(z) = 2\tilde{t}(z/2) + 1 - z\mathrm{e}^{-z/2} - \frac{z^{b+1}\mathrm{e}^{-z}}{(b+1)!}(1-2^{-b}) - \sum_{j=0}^{b} \frac{z^j \mathrm{e}^{-z}}{j!} \left(1 - z2^{-j}\right), \tag{2.3}$$

with the initial conditions being $\tilde{f}(0) = 0$, $\tilde{g}(0) = 0$ *and* $\tilde{t}(0) = 0$.

Proof. Consider $f(z) := \mathrm{e}^z \tilde{f}(z)$. So

$$\begin{aligned} f'(z) &= \sum_{n=0}^{\infty} \frac{\mathbb{E}(N_{n+1})}{n!} z^n = \sum_{n=b}^{\infty} \frac{\mathbb{E}(N_{n+1})}{n!} z^n + \sum_{n=0}^{b-1} \frac{\mathbb{E}(N_{n+1})}{n!} z^n \\ &= \sum_{n=b}^{\infty} 2^{-n+1} \sum_{j=0}^{n} \binom{n}{j} \mathbb{E}(N_j) \frac{z^n}{n!} + \sum_{n=b}^{\infty} \frac{z^n}{n!} + \sum_{n=0}^{b-1} \frac{z^n}{n!} \\ &= \sum_{n=1}^{\infty} 2^{-n+1} \sum_{j=1}^{n} \binom{n}{j} \mathbb{E}(N_j) \frac{z^n}{n!} - \sum_{n=1}^{b-1} 2^{-n+1} \sum_{j=1}^{n} \binom{n}{j} \mathbb{E}(N_j) \frac{z^n}{n!} + \sum_{n=0}^{\infty} \frac{z^n}{n!} \\ &= 2 \sum_{j=1}^{\infty} \mathbb{E}(N_j) \frac{\left(\frac{z}{2}\right)^j}{j!} \sum_{n=j}^{\infty} \frac{\left(\frac{z}{2}\right)^{n-j}}{(n-j)!} - \sum_{n=1}^{b-1} 2^{-n+1} \sum_{j=1}^{n} \binom{n}{j} \frac{z^n}{n!} + \mathrm{e}^z \\ &= 2\mathrm{e}^{\frac{z}{2}} f(z/2) - \sum_{n=1}^{b-1} (2 - 2^{-n+1}) \frac{z^n}{n!} + \mathrm{e}^z. \end{aligned}$$

Then from $\mathrm{e}^{-z} f'(z) = \tilde{f}(z) + \tilde{f}'(z)$, (2.1) will be concluded. Similarly we can get (2.2) and (2.3). □

3. Node-wise path length

We first show that in order to derive asymptotic of $\mathbb{E}(N_n)$ and $\mathbb{E}(X_n)$, it suffices to analyze $\tilde{f}(z)$ and $\tilde{g}(z)$ as $z \to \infty$, respectively. Therefore, we use the theory of analytic de-Poissonization due to [8]. By Definition 1 in [7], a function $\tilde{h}(z)$ is called JS-admissibile if:

(**I**) There exist real numbers α, β such that uniformly for $|\arg(z)| \le \varepsilon$

$$\tilde{h}(z) = \mathcal{O}\Big(|z|^{\alpha} (\log_+ |z|)^{\beta}\Big),$$

where $\log_+ x = \log(1 + x)$.

(**O**) Uniformly for $\varepsilon \le \arg(z) \le \pi$,

$$h(z) := \mathrm{e}^z \tilde{h}(z) = \mathcal{O}\Big(\mathrm{e}^{(1-\varepsilon)|z|}\Big).$$

(Here and throughout the work, ε denotes a small constant whose value might be different from one occurrence to another).
JS-admissibility of given functions is easily checked due to closure properties; see Lemma 2.3 in [7]. Moreover, JS-admissibility of functions $\tilde{f}(z)$ and $\tilde{g}(z)$ which are given by equations of the type (2.1), (2.2) and (2.3) is also easily checked due to the following result has been proved in [7] (Prop. 2.4 in [7]).

Lemma 3.1. *Let $\tilde{h}_1(z)$ and $\tilde{h}_2(z)$ be entire functions with*

$$\tilde{h}_1(z) + \tilde{h}_1'(z) = 2\tilde{h}_1(z/2) + \tilde{h}_2(z)$$

where $\tilde{h}_1(0) = 0$. Then,

$$\tilde{h}_1(z) \text{ is JS-admissible} \qquad \Longleftrightarrow \qquad \tilde{h}_2(z) \text{ is JS-admissible.}$$

Consequently, by Proposition 2.2 in [7], $\mathbb{E}(N_n) \sim \tilde{f}(n)$ and $\mathbb{E}(X_n) \sim \tilde{g}(n)$.

TABLE 1. Numerical approximations to C_b for the first few b.

b	1	2	3	4	5	6
C_b	0	0.82053	0.66699	0.55275	0.46851	0.40514

Theorem 3.2. *Let N_n and X_n be the number nodes and node-wise path length in a random paged DST with page capacity b built from n keys, respectively. As $n \to \infty$, the expectation of N_n and X_n satisfy*

$$\begin{aligned}\mathbb{E}(N_n) &= n\Big(1 - P_b^{[1]}(\log_2 n)\Big) + \mathcal{O}(1),\\ \mathbb{E}(X_n) &= n(\log_2 n)\Big(1 - P_b^{[1]}(\log_2 n)\Big)\\ &\quad + n\Big(\frac{\gamma - 1}{\log 2} + \frac{1}{2} - \alpha + P_b^{[2]}(\log_2 n) - P_b^{[1]}(\log_2 n)\Big)\\ &\quad + (\log_2 n)\Big(1 - P_b^{[3]}(\log_2 n)\Big) + \mathcal{O}(1),\end{aligned}$$

where γ is Euler's constant, $\alpha = \sum_{j\geq 1}(2^j-1)^{-1}$, $P_b^{[1]}$, $P_b^{[2]}$ and $P_b^{[3]}$ are

$$\begin{aligned}P_b^{[1]}(t) &:= \frac{1}{\log 2}\sum_{k\in\mathbb{Z}} \frac{F_b(2+\chi_k)}{\Gamma(2+\chi_k)} \mathrm{e}^{2k\pi i t},\\ F_b(\omega) &:= \sum_{j=1}^{b-1}\int_0^\infty \frac{(2-2^{-j+1})s^{\omega-1}}{Q(-2s)(s+1)^{j+1}}\,\mathrm{d}s,\\ P_b^{[2]}(t) &:= \frac{1}{\log 2}\sum_{k\in\mathbb{Z}\setminus\{0\}} \frac{F(1+\chi_k)}{\Gamma(2+\chi_k)} \mathrm{e}^{2k\pi i t}\\ &\quad + \frac{1}{(\log 2)^2}\sum_{k\in\mathbb{Z}} \frac{F_b(2+\chi_k)\psi(2+\chi_k) - F_b'(2+\chi_k)}{\Gamma(2+\chi_k)} \mathrm{e}^{2k\pi i t},\\ P_b^{[3]}(t) &:= \frac{1}{\log 2}\sum_{k\in\mathbb{Z}} \frac{F_b(2+\chi_k)}{\Gamma(1+\chi_k)} \mathrm{e}^{2k\pi i t},\\ F(\omega) &:= \int_0^\infty \frac{s^{\omega-2}}{Q(-2s)}\,\mathrm{d}s = \frac{Q(2^{\omega-1})}{Q(1)}\Gamma(\omega+1)\Gamma(-\omega),\end{aligned}$$

$\chi_k = 2k\pi i/\log 2$, $Q(z) = \prod_{j\geq 1}(1 - z2^{-j})$ and $\psi = \Gamma'/\Gamma$ is the logarithmic derivative of classic Gamma function. Also $P_b^{[1]}$, $P_b^{[2]}$ and $P_b^{[3]}$ are computable, smooth, 1-periodic functions.

Remark 3.3. The mean value C_b of $1 - P_b^{[1]}(t)$ is given by

$$C_b = \frac{F_b(2)}{\log 2} = \frac{1}{\log 2}\sum_{j=1}^{b-1}\int_0^\infty \frac{(2-2^{-j+1})s}{Q(-2s)(s+1)^{j+1}}\,\mathrm{d}s.$$

Note that when $b = 1$, $C_b = 0$, which is consistent with the fact that $N_n \equiv n$ in this case. As we can see from Table 1, when $b = 5$, for storing n keys we, on the average, need $0.47n$ nodes in a paged DST, meaning that the storage space used by this tree is less than half of that used by a DST.

Proof. We first analyze the mean of N_n by (2.1). Since $\tilde{f}(z)$ is JS-admissible, we may apply Laplace transform to get rid of the differential operator. This yields

$$(s+1)\mathcal{L}[\tilde{f}; s] = 4\mathcal{L}[\tilde{f}; 2s] + \frac{1}{s} - \sum_{j=1}^{b-1}\frac{2-2^{-j+1}}{(s+1)^{j+1}}.$$

Let $\bar{\mathfrak{L}}[\tilde{f};s] := \mathfrak{L}[\tilde{f};s]/Q(-s)$, where $Q(z)=\prod_{j\geq 1}(1-z2^{-j})$. Dividing by $Q(-2s)$ yields

$$\bar{\mathfrak{L}}[\tilde{f};s] = 4\bar{\mathfrak{L}}[\tilde{f};2s] + \frac{1}{Q(-2s)s} - \sum_{j=1}^{b-1}\frac{2-2^{-j+1}}{Q(-2s)(s+1)^{j+1}}. \tag{3.1}$$

Note that we have, by the fact that $N_0=0$ and the proof of Proposition 2.4 in [7], $\tilde{f}(z)=\mathcal{O}(z)$, when $z\to 0^+$ or $z\to\infty$. Then $\mathfrak{L}[\tilde{f};s]=\mathcal{O}(s^{-2})$, if $s\to 0^+$ or $s\to\infty$.

On the other hand, from the fact that $\tilde{f}(z)$ is JS-admissible and well-known growth properties of $Q(-2s)$ (see p. 127 in [7]), we obtain

$$\bar{\mathfrak{L}}[\tilde{f};s] = \begin{cases}\mathcal{O}(s^{-2}), & \text{if} \quad s\to 0^+;\\ \mathcal{O}(s^{-M}), & \text{if} \quad s\to\infty,\end{cases}$$

where $M>0$ is an arbitrary real number. Consequently, the Mellin transform of $\bar{\mathfrak{L}}[\tilde{f};s]$, denoted by $\mathcal{M}[\bar{\mathcal{L}}_f;\omega]$, exists in the half-plane $\Re(\omega)>2$. Thus by applying the Mellin transform to (3.1), we obtain

$$\mathcal{M}[\bar{\mathcal{L}}_f;\omega] = \frac{F(\omega)-F_b(\omega)}{1-2^{2-\omega}},$$

for $\Re(\omega)>2$, where the functions $F(\omega)$ and $F_b(\omega)$ have been defined as in Theorem 3.2 (F_b is an analytic function in the fundamental strip between reals 1 and 2).

From the formula (see p. 316 in [4]),

$$\int_0^\infty \frac{s^{\omega-1}}{Q(-2s)}\,\mathrm{d}s = \frac{\pi Q(2^\omega)}{Q(1)\sin\pi\omega} = \frac{Q(2^\omega)}{Q(1)}\Gamma(\omega+2)\Gamma(-1-\omega),$$

observe that $F(\omega)$ has a simple pole at $\omega=2$. By Proposition 5 in [2], we obtain

$$|F(c+iy)-F_b(c+iy)| = \mathcal{O}\Big(\mathrm{e}^{-(\pi-\varepsilon)|y|}\Big),$$

for large $|y|$ and $c>2$. Then by the calculus of residues, uniformly for $|\arg(s)|\leq\pi-\varepsilon$ and $|s|\to 0$,

$$\bar{\mathfrak{L}}[\tilde{f};s] = \frac{1}{s^2}\left(1-\frac{1}{\log 2}\sum_{k\in\mathbb{Z}}F_b(2+\chi_k)s^{-\chi_k}\right)+\mathcal{O}(|s|^{-1}),$$

where $\chi_k = 2k\pi i/\log 2$. Since $Q(-s)=1+\mathcal{O}(|s|)$, $(|s|\sim 0)$, the above in turn yields

$$\mathfrak{L}[\tilde{f};s] = \frac{1}{s^2}\left(1-\frac{1}{\log 2}\sum_{k\in\mathbb{Z}}F_b(2+\chi_k)s^{-\chi_k}\right)+\mathcal{O}(|s|^{-1}),$$

again uniformly for $|s|\to 0$ and $|\arg(s)|\leq\pi-\varepsilon$. By Proposition 2.6 in [7], we may apply inverse Laplace transform and obtain

$$\tilde{f}(z) = z\left(1-\frac{1}{\log 2}\sum_{k\in\mathbb{Z}}\frac{F_b(2+\chi_k)}{\Gamma(2+\chi_k)}z^{\chi_k}\right)+\mathcal{O}(1),$$

for $|\arg(z)|\leq\frac{\pi}{2}-\varepsilon$ and $z\to\infty$. The same asymptotic expansion also holds for $\mathbb{E}(N_n)$ by de-Poissonization and we deduce the first approximation of Theorem 3.2.

For the second approximation of Theorem 3.2, asymptotic expansion of the mean of X_n, we start from (2.2) and proceed by the same method as above. First, observe that the normalized Laplace transform $\bar{\mathcal{L}}[\tilde{g};s] := \mathcal{L}[\tilde{g};s]/Q(-s)$ satisfies

$$\bar{\mathcal{L}}[\tilde{g};s] = 4\bar{\mathcal{L}}[\tilde{g};2s] + 4\bar{\mathcal{L}}[\tilde{f};2s] - \sum_{j=1}^{b-1} \frac{2-2^{-j+1}}{Q(-2s)(s+1)^{j+1}}.$$

Now, from the fact that $\tilde{g}(z)$ is JS-admissible and the growth properties of $Q(-s)$, we obtain suitable polynomial bounds for $\bar{\mathcal{L}}[\tilde{g};s]$ as s tends both to zero and ∞. This ensures the existence of $\mathcal{M}[\bar{\mathcal{L}}_g;\omega]$, the Mellin transform of $\bar{\mathcal{L}}[\tilde{g};s]$, in a non-trivial strip. Then, we may apply Mellin transform and obtain

$$\mathcal{M}[\bar{\mathcal{L}}_g;\omega] = \frac{2^{2-\omega}F(\omega) - F_b(\omega)}{(1-2^{2-\omega})^2} = \frac{F(\omega-1)}{1-2^{2-\omega}} - \frac{F_b(\omega)}{(1-2^{2-\omega})^2},$$

for $\Re(\omega) > 2$. Consequently, we can proceed as for the mean and derive

$$\begin{aligned}\mathcal{L}[\tilde{g};s] = &\left(\frac{1+s}{s^2}\log_2\frac{1}{s}\right)\cdot\left(1-\frac{1}{\log 2}\sum_{k\in\mathbb{Z}} F_b(2+\chi_k)s^{-\chi_k}\right)\\ &+\frac{1}{s^2}\left(\frac{1}{2}-\alpha\right) - \frac{1}{s^2\log 2}\sum_{k\in\mathbb{Z}} F_b(2+\chi_k)s^{-\chi_k}\\ &+\frac{1}{s^2\log 2}\sum_{k\in\mathbb{Z}\setminus\{0\}} F(1+\chi_k)s^{-\chi_k}\\ &-\frac{1}{s^2(\log 2)^2}\sum_{k\in\mathbb{Z}} F_b'(2+\chi_k)s^{-\chi_k} + \mathcal{O}(|s|^{-1}),\end{aligned}$$

uniformly for $|s|\to 0$ and $|\arg(s)|\le \pi-\varepsilon$, where $\alpha = \sum_{j\ge 1}(2^j-1)^{-1}$.

Finally, standard Laplace inversion gives

$$\begin{aligned}\tilde{g}(z) = &(z+1)\log_2 z + z\left(\frac{\gamma-1}{\log 2}+\frac{1}{2}-\alpha\right) - \frac{z\log_2 z}{\log 2}\sum_{k\in\mathbb{Z}}\frac{F_b(2+\chi_k)}{\Gamma(2+\chi_k)}z^{\chi_k}\\ &-\frac{z}{\log 2}\sum_{k\in\mathbb{Z}}\frac{F_b(2+\chi_k)}{\Gamma(2+\chi_k)}z^{\chi_k} + \frac{z}{\log 2}\sum_{k\in\mathbb{Z}\setminus\{0\}}\frac{F(1+\chi_k)}{\Gamma(2+\chi_k)}z^{\chi_k}\\ &+\frac{z}{(\log 2)^2}\sum_{k\in\mathbb{Z}}\frac{F_b(2+\chi_k)\psi(2+\chi_k) - F_b'(2+\chi_k)}{\Gamma(2+\chi_k)}z^{\chi_k}\\ &-\frac{\log_2 z}{\log 2}\sum_{k\in\mathbb{Z}}\frac{F_b(2+\chi_k)}{\Gamma(1+\chi_k)}z^{\chi_k} + \mathcal{O}(1),\end{aligned}$$

for $|\arg(z)| \le \frac{\pi}{2}-\varepsilon$ and $z\to\infty$, where γ is Euler's constant and $\psi = \Gamma'/\Gamma$ is the logarithmic derivative of Gamma function. The same then holds for $\mathbb{E}(X_n)$ as well by de-Poissonization and the proof will be completed. □

4. 2-PROTECTED NODES

The first paper which studied the number of 2-protected nodes in DSTs was by Du and Prodinger [1], where an asymptotic expansion of the mean was derived by Rice's method. Recently, Fuchs, Lee and Yu [5] show that

TABLE 2. Numerical approximations to C_b^* for the first few b.

b	1	2	3	4	5	6
C_b^*	0.30708	0.25588	0.22272	0.19666	0.17522	0.15740

the previous result and extensions such as the variance and limit laws can be derived in a systematic way. Here, *via* Poisson–Laplace–Mellin method, we obtain the asymptotic expectation of the number of 2-protected nodes for paged DSTs, an important version of DST.

Theorem 4.1. *Let Y_n be the number of 2-protected nodes in a random paged DST with page capacity b built from n keys. As $n \to \infty$, the expectation of Y_n satisfy*

$$\mathbb{E}(Y_n) = n\Big(1 - P_b(\log_2 n)\Big) + \mathcal{O}(1),$$

where P_b is computable, smooth, 1-periodic function and defined by

$$P_b(t) = \frac{1}{\log 2}\sum_{k\in\mathbb{Z}} \frac{G_b(2+\chi_k)}{\Gamma(2+\chi_k)} \mathrm{e}^{2k\pi i t}.$$

Here $\chi_k := 2k\pi i/\log 2$ and for $\Re(\omega) > 2$,

$$G_b(\omega) := \int_0^\infty \frac{s^{\omega-1}}{Q(-2s)}\left(\frac{1}{(s+1/2)^2} + \frac{1-2^{-b}}{(s+1)^{b+2}} + \sum_{j=0}^{b} \frac{s+1-(j+1)2^{-j}}{(s+1)^{j+2}}\right)\mathrm{d}s,$$

where $Q(z) := \prod_{j\ge 1}(1 - z2^{-j})$.

Remark 4.2. The mean value C_b^* of $1 - P_b(t)$ is given by $C_b^* = 1 - \frac{G_b(2)}{\log 2}$. Note that for the random DST case, when $b = 1$, we have $C_b^* = 1 - \frac{G_b(2)}{\log 2} \approx 0.30708$. This matches the value of $\frac{1}{\log 2} \cdot G_1^{(D)}(2)$ in [5]. By Table 2, we see that when $b = 2$, for storing n keys, there are, on the average, $0.25n$ 2-protected nodes in a random paged DST, meaning that asymptotically a proportion of $\frac{1}{4}$ of the nodes is 2-protected. The storage utilization is thus not very bad.

Proof. We first apply the Laplace transform to (2.3) to get rid of the differential operator. This yields the following functional equation

$$(s+1)\mathcal{L}[\tilde{t}; s] = 4\mathcal{L}[\tilde{t}; 2s] + \frac{1}{s} - \frac{1}{(s+1/2)^2} - \frac{1-2^{-b}}{(s+1)^{b+2}} - \sum_{j=0}^{b} \frac{s+1-(j+1)2^{-j}}{(s+1)^{j+2}}.$$

Set $\bar{\mathcal{L}}[\tilde{t}; s] := \mathcal{L}[\tilde{t}; s]/Q(-s)$. Dividing the above equation by $Q(-2s)$ yields

$$\bar{\mathcal{L}}[\tilde{t}; s] = 4\bar{\mathcal{L}}[\tilde{t}; 2s] + \frac{1}{Q(-2s)}\left(\frac{1}{s} - \frac{1}{(s+1/2)^2} - \frac{1-2^{-b}}{(s+1)^{b+2}} - \sum_{j=0}^{b} \frac{s+1-(j+1)2^{-j}}{(s+1)^{j+2}}\right). \tag{4.1}$$

The next step is to derive an asymptotic expansion of $\bar{\mathcal{L}}[\tilde{t}; s]$ by a standard application of Mellin transform; see [3]. Then, note that from [7], we know that

$$\bar{\mathcal{L}}[\tilde{t}; s] = \begin{cases} \mathcal{O}(s^{-2}), & \text{if} \quad s \to 0^+; \\ \mathcal{O}(s^{-M}), & \text{if} \quad s \to \infty, \end{cases}$$

uniformly for s with $|\arg(s)| \le \pi - \varepsilon$, where $M > 0$ is an arbitrary large constant and $\varepsilon > 0$ is an arbitrary small constant. Consequently, the Mellin transform of $\bar{\mathcal{L}}[\tilde{t}; s]$ which we denoted by $\mathcal{M}[\bar{\mathcal{L}}_t; \omega]$, exists in the half-plane $\Re(\omega) > 2$. Moreover, by Proposition 5 in [3], we have, as $|y| \to \infty$,

$$\mathcal{M}[\bar{\mathcal{L}}_t; c + iy] = \mathcal{O}\big(\mathrm{e}^{-(\pi-\varepsilon)|y|}\big),$$

for all $c \in \mathbb{R}$ contained in the fundamental strip. Finally, by applying the Mellin transform to (4.1),

$$\mathcal{M}[\bar{\mathcal{L}}_t; \omega] = \frac{1}{1 - 2^{2-\omega}} \int_0^\infty \frac{s^{\omega-2}}{Q(-2s)} \mathrm{d}s - \frac{G_b(\omega)}{1 - 2^{2-\omega}},$$

for $\Re(\omega) > 0$, where the function G_b is defined as in Theorem 4.1.

From the formula $\int_0^\infty \frac{s^{\omega-1}}{Q(-2s)} \mathrm{d}s = \frac{\pi Q(2^\omega)}{Q(1)\sin \pi\omega} = \frac{Q(2^\omega)}{Q(1)} \Gamma(\omega+2)\Gamma(-1-\omega)$, observe that the first term of the above Mellin transform $\frac{1}{1-2^{2-\omega}} \int_0^\infty \frac{s^{\omega-2}}{Q(-2s)} \mathrm{d}s$ has a simple pole at $\omega = 2$. The second term $\frac{G_b(\omega)}{1-2^{2-\omega}}$ has simple poles at $\omega = 2 + \chi_k$ where $\chi_k = 2k\pi i/\log 2$ for $k \in \mathbb{Z}$. Then by inverse Mellin transform and the calculus of residues

$$\bar{\mathcal{L}}[\tilde{t}; s] = \frac{1}{s^2}\left(1 - \frac{1}{\log 2} \sum_{k \in \mathbb{Z}} G_b(2 + \chi_k) s^{-\chi_k}\right) + \mathcal{O}(|s|^{-1}),$$

uniformly as $|s| \to 0$ and $|\arg(s)| \le \pi - \varepsilon$. Moreover, due to $Q(-s) = 1 + \mathcal{O}(|s|)$ for $|s| \to 0$, (see p. 127 in [7]), the same asymptotic expansion holds for $\mathcal{L}[\tilde{t}; s]$ as well.

Next, we apply inverse Laplace transform. More precisely, we use Proposition 2.6 of [7] and obtain

$$\tilde{t}(z) = z\left(1 - \frac{1}{\log 2} \sum_{k \in \mathbb{Z}} \frac{G_b(2 + \chi_k)}{\Gamma(2 + \chi_k)} z^{\chi_k}\right) + \mathcal{O}(1),$$

uniformly as $z \to \infty$ and $|\arg(z)| \le \frac{\pi}{2} - \varepsilon$.

The final step is to use de-Poissonization in order to get an asymptotic expansion of $\mathbb{E}(Y_n)$ from that of $\tilde{t}(z)$; see [8] and Section 2.3 in [7]. From the closure properties proved in this section (see Lem. 2.3) and Proposition 2.4, we obtain that $\tilde{t}(z)$ is JS-admissible. Hence, by Proposition 2.2 in [7], our claimed result for 2-protected nodes in Theorem 4.1 is proved. □

References

[1] R.R.X. Du and H. Prodinger, Notes on Protected Nodes in Digital Search Trees. *Appl. Math. Lett.* **25** (2012) 1025–1028.

[2] P. Flajolet, X. Gourdon and P. Dumas, Mellin Transforms and Asymptotics: Harmonic Sums. *Theor. Comput. Sci.* **144** (1995) 3–58.

[3] P. Flajolet, X. Gourdon and C. Martinez, Patterns in Random Binary Search Trees. *Random Struct. Algorithms.* **11** (1997) 223–244.

[4] P. Flajolet and B. Richmond, Generalized Digital Trees and Their Difference-Differential Equations. *Random Struct. Algorithms* **3** (1992) 305–320.

[5] M. Fuchs, C.K. Lee and G.R. Yu, On 2-Protected Nodes in Random Digital Trees. *Theor. Comput. Sci.* **622** (2016) 111–122.

[6] M. Hoshi and P. Flajolet, Page Usage in a Quadtree Index. *BIT* **32** (1992) 384–402.

[7] H.K. Hwang, M. Fuchs and V. Zacharovas, Asymptotic Variance of Random Symmetric Digital Search Trees. *Discrete Math. Theor. Comput. Sci.* **12** (2010) 103–166.

[8] P. Jacquet and W. Szpankowski, Analytical De-Poissonization and Its Applications. *Theor. Comput. Sci.* **201** (1998) 1–62.

[9] D.E. Knuth, *The Art of Computer Programming in: Sorting and Searching.* Addison Wesley (1998).

[10] C. Martinez, A. Panholzer and H. Prodinger, On the Number of Descendants and Ascendants in Random Search Trees. *Electron. J. Combin.* **5** (1998) #R20.

[11] J.S. Vitter and P. Flajolet, Average-case analysis of algorithms and data structures. *Handb. Theor. Comput. Sci.* (1990).

Communicated by P. Boldi.
Received February 1, 2016. Accepted January 10, 2017.