Theoretical Computer Science 622 (2016) 111–122

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

On 2-protected nodes in random digital trees

M. Fuchs [a,*], C.-K. Lee [b], G.-R. Yu [c]

[a] Department of Applied Mathematics, National Chiao Tung University, Hsinchu, 300, Taiwan
[b] Institute of Information Sciences, Academia Sinica, Taipei, 11529, Taiwan
[c] Institute for Discrete Mathematics and Geometry, Technical University of Vienna, 1040 Vienna, Austria

ARTICLE INFO

Article history:
Received 29 September 2015
Received in revised form 28 January 2016
Accepted 8 February 2016
Available online 12 February 2016
Communicated by H. Prodinger

Keywords:
Data structures
Digital trees
Analytic combinatorics
Moments
Limit theorems
Abel summability

ABSTRACT

In this paper, we consider the number of 2-protected nodes in random digital trees. Results for the mean and variance of this number for tries have been obtained by Gaither et al. (2012) [11] and Gaither and Ward (2013) [10] and for the mean in digital search trees by Du and Prodinger (2012) [5]. In this short note, we show that these previous results and extensions such as the variance in digital search trees and limit laws in both cases can be derived in a systematic way by recent approaches of Fuchs et al. (2012; 2014) [8, 15] and Fuchs and Lee (2014) [9]. Interestingly, the results for the moments we obtain by our approach are quite different from the previous ones and contain divergent series which have values by appealing to the theory of Abel summability. We also show that our tools apply to PATRICIA tries, for which the number of 2-protected nodes has not been investigated so far.

1. Introduction

2-protected nodes in rooted trees are nodes with a distance of at least 2 to every leave (or in other words, nodes which are neither leaves nor parents of leaves). They were introduced by Cheon and Shapiro in [3] as an efficiency measure of organizational schemes. Other applications such as in social networks models have been discussed by Gaither and Ward in [10]. Apart from the practical motivation, the study of 2-protected nodes and their obvious generalization to k-protected nodes is also interesting from a theoretical point of view. More precisely, k-protected nodes, when considered as a sequence of k, can be interpreted as a profile describing the tree from the fringe to the root. Many other notions of profiles have been investigated in recent years for many different types of random trees. Studying 2-protected nodes constitutes the first step in the study of such a "protected node profile". That is why they have been investigated for many different random trees by many authors: Cheon and Shapiro [3] (random binary trees and random Motzkin trees); Mansour [20] (random k-ary trees); Bóna [2], Devroye and Janson [4], Holmgren and Janson [12,14], Mahmoud and Ward [18] (random binary search trees); Devroye and Janson [4], Holmgren and Janson [12,14], Mahmoud and Ward [19] (random recursive trees); Holmgren and Janson [13] (random ternary search trees); and Devroye and Janson [4] (simple generated families of random trees).

In this short note, we are interested in the number of 2-protected nodes in the three main families of random digital trees, namely, random tries (invented by de la Briandais), random PATRICIA tries (invented by Morrison) and random digital search trees (invented by Coffman and Eve). Apart from PATRICIA tries which have not been treated before, results on moments for the number of 2-protected nodes for the other two types of random digital trees already exist. Before recalling

* Corresponding author. Tel.: +88635712121 56461.
E-mail address: mfuchs@math.nctu.edu.tw (M. Fuchs).

http://dx.doi.org/10.1016/j.tcs.2016.02.007

these earlier results, we will give the definition of the above three families which are all fundamental data structures in computer science; for more background see Mahmoud [17] or [8,15].

Now, for the definition, assume that n infinite 0-1 sequences are given which are records to be stored in a binary tree. From these records the trie is a binary tree built as follows: if $n = 1$, the only record is stored in the root; if $n > 1$, then the root is an (empty) internal node and all records are distributed to the two subtrees according to whether their first bit is 0 or 1; finally, the subtrees are built recursively according to the same rules, but by considering subsequent bits. PATRICIA tries are built by the same procedure with the only difference that one-way branching is avoided. Finally, digital search trees are also built similarly, but now records may be stored in internal nodes, too. In the sequel, we will equip these tree families with the Bernoulli model which means that bits are i.i.d. Bernoulli random variables with success probability p. We also set $q := 1 - p$ throughout the paper.

We next recall what is known about the number of 2-protected nodes in random digital trees. The first paper which studied this parameter was by Du and Prodinger [5], where an asymptotic expansion of the mean in symmetric digital search trees (i.e., $p = q = 1/2$) was derived. Then, Gaither, Homma, Sellke and Ward [11] proved a similar result for random tries, but for the general case (not only the symmetric case). Moreover, Gaither and Ward in [10] found an asymptotic expansion for the variance of the number of 2-protected nodes in tries and announced a central limit theorem, which was conjectured in their paper. Finally, Devroye and Janson in [4] also announced a (possible) future study of 2-protected nodes in random digital trees based on their method from [4].

The main aim of this paper is to show that all previous results as well as more refined properties for 2-protected nodes in random digital trees can be systematically derived with tools which were developed in a recent series of papers by Fuchs, Hwang and Zacharovas [15,8] and a paper of Fuchs and Lee [9] (these papers were concerned with general frameworks for studying stochastic properties of parameters such as 2-protected nodes in random digital trees; for details see below). Moreover, our tools also apply straightforwardly to PATRICIA tries, which have not been treated before. For mean and variance, we will compare our results with previous results (if already known). Interestingly, the expressions we obtain will be rather different. In particular, the periodic functions in our expressions for tries will considerably differ from those in [10, 11] and will contain divergent series which can be made convergent by appealing to the theory of Abel summability. This is a new phenomena which has not been present in any of the examples studied in [8]. Apart from considering moments, we will also look at limiting distributions and prove (univariate and bivariate) central limit theorems for the number of 2-protected nodes in the three types of random digital trees. For tries, this will confirm the above mentioned conjecture of [10]. In all other cases, our results are new.

We conclude the introduction with a short sketch of the paper. In the next section, we will recall the framework from [8] and [9]. Moreover, we will sketch a similar framework for symmetric digital search trees which is based on [15] and which was recently obtained by Lee in his Ph.D. thesis [16]. In Section 3, we will discuss our results for random tries and PATRICIA tries. Finally, Section 4 will contain corresponding results for symmetric digital search trees. In the proof of all of our results, we will be deliberately brief since (i) as mentioned above our intention is to show that results for 2-protected nodes for random digital trees follow quite straightforwardly from previous studies and (ii) we do not want to repeat things which have appeared in previous works.

Notations Throughout the paper, the number of 2-protected nodes in a random digital tree of size n under the Bernoulli model will be denoted by $X_n^{(\star)}$ with $\star \in \{T, P, D\}$, depending on whether tries, PATRICIA tries, or digital search trees are considered. Moreover, for some function $G(x)$, we will use the notation

$$\mathscr{F}[G](x) := \begin{cases} \dfrac{1}{h} \displaystyle\sum_{k \in \mathbb{Z}\setminus\{0\}} G(-1+\chi_k) e^{2k\pi i x}, & \text{if } \frac{\log p}{\log q} \in \mathbb{Q}; \\ 0, & \text{if } \frac{\log p}{\log q} \notin \mathbb{Q}, \end{cases}$$

where $h = -p\log p - q\log q$ and $\chi_k = 2rk\pi i/\log p$ when $\log p/\log q = r/l$ with $\gcd(r,l) = 1$.

2. Preliminaries

In this section, we are going to recall the results from [8] and [9]. We will state them in a form convenient for the applications below. We first need the following notation.

Definition 1. Let $\tilde{f}(z)$ be an entire function and $\alpha, \gamma \in \mathbb{R}$. Then, we say that $\tilde{f}(z)$ is JS-admissible (named after Jacquet and Szpankowski who did important work on these functions) and write $\tilde{f}(z) \in \mathscr{JS}$ (or more precisely, $\tilde{f}(z) \in \mathscr{JS}_{\alpha,\gamma}$) if for $0 < \phi < \pi/2$ and all $|z| \geq 1$ the following two conditions hold.

(I) Uniformly for $|\arg(z)| \leq \phi$,

$$\tilde{f}(z) = \mathcal{O}\left(|z|^{\alpha} (\log_+ |z|)^{\gamma}\right),$$

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