Protection Number of Recursive Trees *

Zbigniew Gołębiewski† Mateusz Klimczak‡

Abstract

The protection number of a tree is the minimal distance from its root to a leaf. In this paper we are interested in the protection number of a uniformly chosen random recursive tree of size n. Due to different construction of plane oriented and non-plane recursive trees we consider them separately. We use the singularity analysis of derived generating functions to find out the number of relevant trees, which leads us to the probability distribution of the protection number. Our results are also compared to outcomes of computer simulations.

1 Introduction

In a rooted tree, a vertex is said to be k-protected if its minimum distance from a leaf in its subtree is at least k. A tree is said to be k-protected if its root is k-protected. The maximal k such that a vertex v is k-protected is called the protection number of v. The protection number of a tree equals to the protection number of the root of this tree.

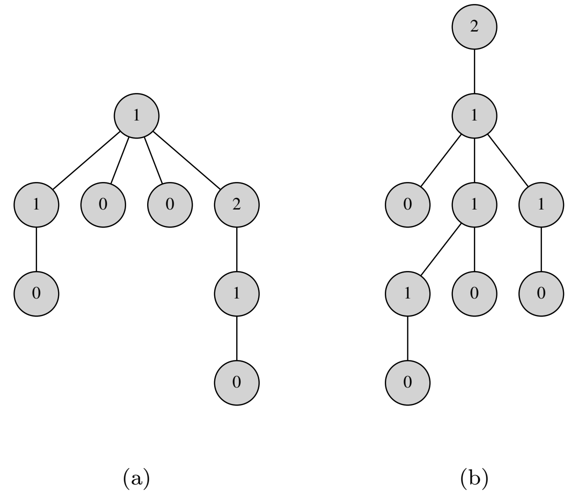

Figure 1: Two examples of trees where vertices holds their protection number.

Studies of the number of 2-protected nodes in trees (the case of unlabelled ordered trees and Motzkin trees) were started by Cheon and Shapiro in [17]. Later Mansour in [22] complemented their results by studying k-ary trees. During the next several years there appeared papers that investigated the number of k-protected nodes (usually for a small values of k) for different models of random trees. In [23] Du and Prodinger analysed the average number of 2-protected nodes in random digital trees, in [25] Mahmoud and Ward presented a central limit theorem, as well as exact moments of all orders, for the number of 2-protected nodes in binary search trees. The case of larger k values in binary search trees were studied by Bona and Pittel in [28]. The number of 2-protected nodes in recursive trees have been studied by Mahmoud and Ward in [27]. In [26] Devroye and Janson presented an unified approach to study 2-protected nodes in various classes or random trees by putting them in the general context of fringe subtrees introduced by Aldous in [5]. Advanced probabilistic methods were used by Holmgren and Janson in [24] where they obtained a normal limit law for the number of k-protected nodes in a binary search tree or a random recursive tree.

Copenhaver in [29] found the proportion of vertices which are k-protected in unlabelled rooted plane trees. Moreover, he also found that, as n goes to infinity, the expected value of the protection number of a random vertex in a tree of size n approaches 0.727649, and the expected value of the protection number of the root of a tree of size n approaches 1.62297. These results were extended by Heuberger and Prodinger in [30]. We should mention that the protection number of the root of a tree was studied before under different names: the minimal fill-up level or the saturation level. Devroye in [7] showed that in a random PATRICIA tree of n nodes the minimal fill-up level F_n satisfies: $\frac{F_n - \log_2 n}{\log_2 \log n} \to -1$ almost surely. The saturation level $\overline{H}_n$ in d-ary recursive trees of size n was studied by Drmota in [12, 19]. It was shown that $\mathbb{E}\left[\overline{H}_n\right] \sim \alpha_d \log n$, where

*The work was supported by Wroclaw University of Science and Technology grant number 0401/0052/18.

†Department of Computer Science, Faculty of Fundamental Problems of Technology, Wroclaw University of Science and Technology.

‡Department of Computer Science, Faculty of Fundamental Problems of Technology, Wroclaw University of Science and Technology.

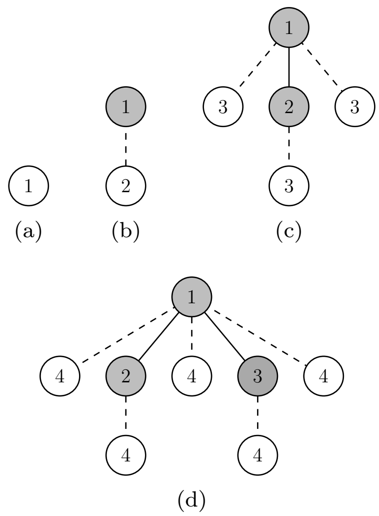

Figure 2: Example of the first four steps of generating plane oriented recursive trees. Gray nodes are already placed and white ones indicates possible locations for attaching the next node.

the constant α_d is the smaller solution of the equation $\alpha \log\left(\frac{de}{(d-1)\alpha}\right) = \frac{1}{d-1}$.

Our main focus is to study the protection number of a random recursive tree (i.e. the protection number of the root of a random recursive tree). In this extended abstract we present a full analysis of the protection number of plane oriented recursive trees and we state some results in case of non-plane recursive trees.

A recursive tree is a rooted labelled tree such that the root is labelled by 1 and the labels of all children of any node u are different and larger than the label of u. For convenience we assume that labels in a recursive tree with n nodes form a set of numbers from 1 to n. While considering recursive trees, we may take the order of node children into account or we may not. Therefore we distinguish plane oriented recursive trees (abbreviated as PORTs) and non-plane recursive trees.

Plane oriented recursive trees have been introduced by Szymański in [3]. They can be also seen as a result of the following process. We begin with a single node with label 1. Then we can attach the node with label 2 only in one way - as a child of 1. The node with label 3 can be attached as a child of 1 (left or right sibling of 2) or as a child of 2. We continue this process in similar manner to get a tree with n nodes. In each step we choose a place for a new node uniformly at random. The first four steps of this process are shown in Figure 2.

From the above description we see that if a node u has out-degree d, then we have $d+1$ ways to attach a new node to u. Therefore, having $n-1$ nodes the number of all possibilities to add the n-th node equals $2n-3$. Consequently the number of all plane oriented recursive trees with n nodes is $1 \cdot 3 \cdots (2n-3) = (2n-3)!!$. The natural probability distribution on PORTs of size n is to assume that each of these $(2n-3)!!$ trees is equally likely. This probability distribution is also obtained from the above described evolution process.

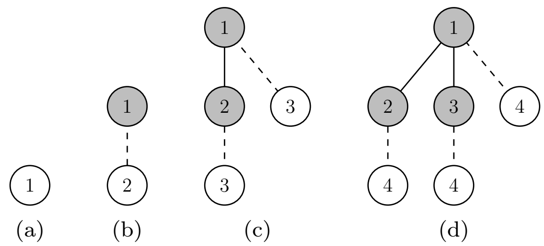

Figure 3: Example of the first four steps of generating non-plane recursive trees. Gray nodes are already placed and white ones indicates possible locations for attaching the next node.

On the other hand, non-plane recursive trees can be seen as the result of the following evolution process. Suppose that the process starts with a node carrying the label 1. This node will be the root of the tree. Then we attach a node with label 2 to the root. The node with label 3 can be attached as a child of the root or to the node with label 2. We continue this process in the similar manner to get a tree with n nodes. After attaching k nodes, we attach the node with label $k+1$ to one of the existing nodes chosen uniformly at random. The first four steps of this process are shown in Figure 3.

For each existing node we have one way to attach a new one. Therefore, having $k-1$ nodes the number of all possibilities to add k-th node simply equals $k-1$. Consequently, the number of all non-plane recursive trees with n nodes is $(n-1)!$.

A vast number of papers have been written about recursive trees and their properties. Let us just refer to the article by Bergeron, Flajolet and Salvy [6] that was a starting point for studying recursive trees with use of analytic combinatorics tools. Properties like insertion depth, path length [8, 11], height [4, 10, 16, 18], degree distribution [2, 14] and profile [2, 13, 15] of recursive trees were studied over the years. These trees have found plenty of applications as data structures in the computer science, as models in biology, and as representations of permutations, to name a few [9, 19, 21].

Paper Organization In Section 2 we give a detailed analysis of the probability distribution of the protection number of a random plane oriented recursive tree. Section 3 consist of the results about the probability distribution of the protection number of a random non-plane recursive tree. We finish the paper with some concluding remarks and future works directions in Section 4.

2 Protection Number of Plane Oriented Recursive Trees

Let $\mathcal{L}_n$ be a set of plane oriented recursive trees with n nodes. Let a random variable X_n denote the protection number of randomly chosen PORT with n nodes (we assume uniform distribution on the elements of the set $\mathcal{L}_n$).

Let $\mathcal{L}_{n,k} \subseteq \mathcal{L}_n$ be a set of all k-protected PORTs on n nodes and let $l_{n,k} = |\mathcal{L}_{n,k}|$ denote its cardinality. Observe that $\mathcal{L}_n = \mathcal{L}_{n,0} \supseteq \mathcal{L}_{n,1} \supseteq \mathcal{L}_{n,2} \supseteq \ldots$. Let us introduce the labelled combinatorial classes $(\mathcal{L}_{\geq k}, |\cdot|)$, where $\mathcal{L}_{\geq k} = \bigcup_{n\geq 0} \mathcal{L}_{n,k}$ is the set of all k-protected plane oriented recursive trees and the size function $|\cdot|$ counts the number of nodes in a tree. Let $L_k(z)$ be the exponential generating function of the class $\mathcal{L}_{\geq k}$:

$$L_k(z) = \sum_{n\geq 0} |\mathcal{L}_{n,k}| \frac{z^n}{n!} = \sum_{n\geq 0} l_{n,k} \frac{z^n}{n!}.$$

We investigate those functions in order to get meaningful information about the distribution of X_n. We observe that for any $k \geq 1$, each k-protected plane oriented recursive tree is a pair consisting of the label of the root and a non-empty sequence of subtrees whose roots are $(k-1)$-protected. There is also an order constraint – the smallest label belongs to the root. From this specification we have the following symbolic equation (for more about symbolic methods see [20]):

$$\mathcal{L}_{\geq k} = \mathcal{Z}^{\square} \star \textsc{Seq}_{\geq 1}(\mathcal{L}_{\geq k-1}).$$

Notice that $\sum_{n\geq 0}(2n-3)!!\frac{z^n}{n!} = 1-\sqrt{1-2z}$. This leads us to the following recursive system of equations

$$(2.1) \qquad \begin{cases} L_k(z) = \int_0^z \frac{1}{1-L_{k-1}(t)}dt - z & \text{for } k \geq 1, \\ L_0(z) = 1 - \sqrt{1-2z}. \end{cases}$$

Using standard integration by substitution we are able to calculate $L_1(z) = 1 - \sqrt{1-2z} - z = L_0(z) - z$. This result is not surprising, because the only tree which is 0-protected and is not 1-protected is a single node. The next step is to calculate $L_2(z)$. It is slightly more difficult, but still doable and we obtain:

$$L_2(z) = \frac{1}{2}\Big(-\ln 4 + \left(2+\sqrt{2}\right)\ln\left(1+\sqrt{2}-\sqrt{1-2z}\right) - \left(-2+\sqrt{2}\right)\ln\left(-1+\sqrt{2}+\sqrt{1-2z}\right)\Big) - z.$$

The first coefficients of $L_2(z)$ are $0, 0, 0, 1, 3, 21, 165, 1665, 19845, 275625, 4363065, \ldots$. Surprisingly this sequence is not present in the OEIS database.

A precise calculation of $L_k(z)$ for $k \geq 3$ seems to be hopeless, so we turn to finding an asymptotic of $L_k(z)$ near its singularity. In order to do this, we will make a few important observations. In the following lemma we determine the radius of convergence of functions L_k for any $k \geq 0$. It is a crucial observation that already dictates the exponential growth of the coefficients of $L_k(z)$.

Lemma 2.1. *For any $k \geq 0$ the function $L_k(z)$ has its nearest to origin singularity at $\rho = \frac{1}{2}$.*

Proof. Let ρ_k be the convergence radius of $L_k(z)$. We will show that for all $k \in \mathbb{N}$: $\rho_k = \rho = \frac{1}{2}$. Since $L_0(z) = 1 - \sqrt{1-2z}$ we know that the convergence radius of $L_0(z)$ is $\rho_0 = \frac{1}{2}$.

Now, we argue that the convergence radius does not increase with k. The function $L_{k+1}(z)$ has the same singularities as $\int_0^z \frac{dt}{1-L_k(t)}$. The integration does not change the convergence radius, so singularities come only from $\frac{1}{1-L_k(t)}$. Thus we certainly have some singularities with modulus ρ_k and in z_k such that $L_k(z_k) = 1$. This means that the convergence radius ρ_{k+1} cannot be larger than ρ_k.

Next, we show that the convergence radius does not decrease with k. The coefficients $l_{n,k} = [z^n]L_k(z)$ form a decreasing sequence with respect to k because, as we already noted, $\mathcal{L}_{n,k} \supseteq \mathcal{L}_{n,k+1}$. From [20, Thm. IV.7] we know that series coefficient $[z^n]L_k(z)$ is of exponential order ρ_k^{-n}. So, the inequality $[z^n]L_k(z) \geq [z^n]L_{k+1}(z)$ implies $\rho_k^{-n} \geq \rho_{k+1}^{-n}$, which yields $\rho_k \leq \rho_{k+1}$.

From *Pringsheim's Theorem* [20, Thm. IV.6] we know that the nearest singularity lies on $\mathbb{R}_+$.

Putting everything together yields that for all $k \geq 1$ $\rho_{k-1} = \rho_k$, so for all $k \in \mathbb{N} : \rho_k = \rho_0 = \frac{1}{2}$.

The next auxiliary lemma will be used in several places in the paper in order to get a precise information about the $L_k(z)$ functions, especially where their singularities come from etc.

Lemma 2.2. *For any $k \geq 1$ and any complex number $z = re^{i\theta}$ with $0 \leq r \leq \frac{1}{2}$ we have $|L_k(z)| < 1$.*

Proof. Notice that $|z^n| = r^n$ and by the triangle inequality for countable number of terms we get

$$|L_k(z)| = \left|\sum_{n\geq 0} l_{n,k}\frac{z^n}{n!}\right| \leq \sum_{n\geq 0} l_{n,k}\frac{|z^n|}{n!} = \sum_{n\geq 0} l_{n,k}\frac{r^n}{n!} = L_k(r).$$

On the other hand we know that the $l_{n,k}$ form a decreasing sequence with respect to k because $L_{n,k} \supseteq L_{n,k+1}$. So the fact that for all $n \in \mathbb{N} : l_{n,k} \geq l_{n,k+1}$ implies that for all $x \in \mathbb{R}_+ : L_k(x) \geq L_{k+1}(x)$. In general we have

$$(\forall x \in \mathbb{R}_+) \quad L_0(x) \geq L_1(x) \geq L_2(x) \geq \dots$$

In particular, from the exact formulas for $L_0(z), L_1(z)$ and $L_2(z)$ for $x = \frac{1}{2}$ we get

$$\begin{array}{ccccccc} L_0\left(\frac{1}{2}\right) & \geq & L_1\left(\frac{1}{2}\right) & \geq & L_2\left(\frac{1}{2}\right) & \geq & \dots \\ \parallel & & \parallel & & \parallel & & \\ 1.0 & \geq & 0.5 & \geq & 0.053\dots & \geq & \dots \end{array}$$

Therefore for any $k \geq 1$ and for any complex number $z = re^{i\theta}$ with $0 \leq r \leq \frac{1}{2}$ we have

$$|L_k(z)| \leq L_k(r) \leq L_1(r) \leq L_1\left(\frac{1}{2}\right) = \frac{1}{2} < 1. \tag{2.2}$$

The next lemma gives us the precise location of the dominant singularity of the $L_k(z)$ functions.

LEMMA 2.3. *For any $k \geq 0$ the function $L_k(z)$ has no complex singularities with modulus $r = \frac{1}{2}$.*

Proof. We know that a function is analytic in a point z_0 if and only if it is complex-differentiable in z_0 (from [20, Thm. IV.1]). From the recursive specification of $L_k(z)$ (see Equation (2.1)) we also know the formula for the complex derivative of $L_{k+1}(z)$ for $k \geq 0$:

$$L'_{k+1}(z) = \frac{1}{1 - L_k(z)} - 1.$$

Thus $L_{k+1}(z)$ has the same singularities as L_k and additionally in those points, where $L_k(z)$ is equal to 1. We are interested only in singularities with modulus $r = \frac{1}{2}$. The function $L_1(z)$ has its singularities in those points where $L_0(z)$ has and in those points where $L_0(z) = 1$. We see that it happens only in $\rho = \frac{1}{2}$. Moreover, by Lemma 2.2 we notice that for $k \geq 1$ the function $L_{k+1}(z)$ only inherits singularities from $L_k(z)$, because $L_k(z)$ is less than 1 in the circle with radius $\frac{1}{2}$. Therefore, for all $k \geq 1$ the function $L_k(z)$ has only one singularity with modulus $r = \frac{1}{2}$ and it is $\rho = \frac{1}{2}$.

Now, we are in the position to give one of the main results of this paper. As we observe in the previous lemmas, it turns out that all $L_k(z)$ functions have their dominant singularities in the same place $\rho = \frac{1}{2}$. Now, we have to determine the nature of those singularities which will give us information about the subexponential factors of the coefficients of the $L_k(z)$ functions.

THEOREM 2.1. *For any $k \geq 1$ we have $L_k(z) = b_k(z) + c_k(1-2z)^{k-\frac{1}{2}} + O\left((1-2z)^k\right)$ as $z \to \frac{1}{2}$ where $b_k(z)$ is a polynomial of degree at most $k-1$ and c_k is some constant.*

Proof. We will prove this theorem by induction. For $k = 1$ we have

$$\begin{aligned} L_1(z) &= 1 - \sqrt{1-2z} - z \\ &= \underbrace{\frac{1}{2}}_{b_1(z)} + \underbrace{(-1)}_{c_1} \cdot (1-2z)^{\frac{1}{2}} + \underbrace{\frac{1}{2}(1-2z)}_{O(1-2z)}, \end{aligned}$$

so we see that in this case theorem holds. Assume that theorem holds for some $k \geq 1$.

To obtain $L_{k+1}(z)$ we have to integrate $\frac{1}{1-L_k(z)}$ and subtract z at the end. Now we take a closer look at the function $\frac{1}{1-L_k(z)}$. By Lemma 2.1 we know that $L_k(z)$ has its nearest to the origin singularity in $\frac{1}{2}$. Moreover, by Lemma 2.2 we have $|L_k(z)| < 1$ for $|z| \leq \frac{1}{2}$, so the function $\frac{1}{1-L_k(z)}$ inherits the nearest to origin singularity from $L_k(z)$. Thus, we can expand $\frac{1}{1-L_k(z)}$ near its singularity $\rho = \frac{1}{2}$ and get

$$\frac{1}{1 - L_k(z)} = h_k(z) + h_{k,\bullet}(1-2z)^{k-\frac{1}{2}} + O\left((1-2z)^k\right),$$

where $h_k(z) = h_{k,0} + h_{k,1}(1-2z) + \ldots + h_{k,k-1}(1-2z)^{k-1}$ is a polynomial of degree at most $k-1$. Moreover, having that $b_k(z) = b_{k,0} + b_{k,1}(1-2z) + \cdots + b_{k,k-1}(1-2z)^{k-1}$ we obtain $h_{k,0} = \frac{1}{1-b_{k,0}}$ and $h_{k,\bullet} = \frac{c_k}{(1-b_{k,0})^2}$. Then by *Singular Integration Theorem* (see [20, Thm. VI.9]) and subtracting z we get

$$\begin{aligned} L_{k+1}(z) = &\underbrace{\int_0^{\frac{1}{2}} \frac{dt}{1 - L_k(t)} - z - \frac{1}{2}\sum_{j=0}^{k-1} \frac{h_{k,j}}{j+1}(1-2z)^{j+1}}_{b_{k+1}(z)} \\ &+ \underbrace{\frac{-h_{k,\bullet}}{2k+1}}_{c_{k+1}}(1-2z)^{k+\frac{1}{2}} + O\left((1-2z)^{k+1}\right). \end{aligned}$$

The integral $\int_0^{\frac{1}{2}} \frac{dt}{1-L_k(t)}$ is finite, because $L_k(z) \leq M < 1$ is bounded on the closed interval $[0, \frac{1}{2}]$ and thus the function $\frac{1}{1-L_k(z)}$ is also bounded by $\frac{1}{1-M}$ in that interval. Concluding, by the principle of mathematical induction, the theorem holds for all $k \geq 1$.

Observe that for the growth of the series coefficients, the most important term in the obtained singular expansion of $L_k(z)$ is $c_k(1-2z)^{k-\frac{1}{2}}$. Therefore the nature of the dominant singularity, and consequently subexponential factor of the n'th coefficient of $L_k(z)$ function, depends on k. It is also worth noticing that in case of Catalan trees (i.e. rooted unlabelled plane trees) it is not the case. In [30] Heuberger and Prodinger showed that the main term in the singular expansion of the generating function of the number of k-protected unlabelled plane trees equals $\frac{-9\cdot 4^k}{2\cdot 4^{2k}+8\cdot 4^k+8}(1-4z)^{\frac{1}{2}}$ for any $k \geq 0$. So, the exponent equals $\frac{1}{2}$ for any $k \geq 0$ and only the constant depends on k.

Remark. From the proof of Theorem 2.1 it is possible to obtain an upper bound for the constant c_k. Notice that we have the following recurrence equation

$$c_{k+1} = \frac{-h_{k,\bullet}}{2k+1} = \frac{-c_k}{(2k+1)(1-b_{k,0})^2}.$$

Since $b_{k,0} = L_k(\frac{1}{2})$ and from Inequality (2.2) we know that for all $k \geq 1$: $L_k(\frac{1}{2}) \leq \frac{1}{2}$, we have that $\frac{1}{(1-b_{k,0})^2} \leq 4$. Therefore

$$c_{k+1} \leq \frac{-4c_k}{2k+1},$$

and since $c_1 = -1$ we get the following bound

$$c_k \leq \frac{(-1)^k 4^{k-1}}{(2k-1)!!} = \frac{(-1)^k 2^{3k-2} k!}{(2k)!}. \tag{2.3}$$

Now, using *Transfer Theorem* (see [20, Thm. VI.1]), double factorial and Gamma functions properties, we easily get the following.

Corollary 2.1. *Asymptotic number of k-protected plane oriented recursive trees of size n equals*

$$l_{n,k} = n!\,[z^n]\,L_k(z) = n!\;2^n\left(\frac{d_k}{n^{k+\frac{1}{2}}} + \mathrm{O}\left(\frac{1}{n^{k+1}}\right)\right)$$

as $n \to \infty$ where $d_k = \frac{c_k(-1)^k(2k)!}{4^k k!\sqrt{\pi}}$.

In particular we can calculate the "expected result" that $l_{n,1} \sim \frac{n!2^n}{2\sqrt{\pi n^3}}$ (with $d_1 = \frac{1}{2\sqrt{\pi}}$) as well as the result for the number of 2-protected PORTs of size n: $l_{n,2} \sim \frac{n!2^n}{\sqrt{\pi n^5}}$ (with $d_2 = \frac{1}{\sqrt{\pi}}$) as $n \to \infty$.

Additionally, with use of Inequality (2.3) we have for all $k \geq 1$ the following upper bound:

$$d_k \leq \frac{2^{k-2}}{\sqrt{\pi}}. \tag{2.4}$$

With Theorem 2.1 in hand we can derive interesting results concerning the probability distribution of X_n the protection number of randomly chosen PORT of size n.

Theorem 2.2. *For $k \geq 1$, probability that a uniformly chosen plane oriented recursive tree with n nodes is k protected is*

$$\mathbb{P}(X_n \geq k) = 2d_k\sqrt{\pi}n^{1-k} + \mathrm{O}\left(n^{\frac{1}{2}-k}\right) \quad \textit{as } n \to \infty.$$

Proof. Observe that having uniform distribution among the trees of size n we have

$$\mathbb{P}(X_n \geq k) = \frac{|\mathcal{L}_{n,k}|}{|\mathcal{L}_n|} = \frac{n![z^n]L_k(z)}{n![z^n]L_0(z)} = \frac{l_{n,k}}{l_{n,0}}.$$

Using the asymptotics of $l_{n,k}$ from Corollary 2.1 we get the desired result.

Remark. Since all but single node tree are 1-protected we obviously have the following exact result:

$$\mathbb{P}(X_n \geq 1) = \begin{cases} 0 & \text{if } n \leq 1, \\ 1 & \text{otherwise.} \end{cases} \tag{2.5}$$

Corollary 2.2. *Mean and variance of the protection number of uniformly chosen random plane oriented recursive tree with n nodes equals*

$$\mathbb{E}[X_n] = 1 + \frac{2}{n} + \mathrm{O}\left(\frac{1}{n^{\frac{3}{2}}}\right)$$
$$\mathbb{V}[X_n] = \frac{2}{n} - \frac{4}{n^2} + \mathrm{O}\left(\frac{1}{n^{\frac{5}{2}}}\right).$$

Proof. Since X_n is non-negative and integer valued random variable and we already know that $d_2 = \frac{1}{\sqrt{\pi}}$, we have

$$\begin{aligned}\mathbb{E}[X_n] &= \sum_{k\geq 1}\mathbb{P}(X_n \geq k) \\ &= 1 + \frac{2}{n} + \mathrm{O}\left(\frac{1}{n^{\frac{3}{2}}}\right) + \sum_{k\geq 3}\mathbb{P}(X_n \geq k).\end{aligned}$$

Now we will show that $\sum_{k\geq 3}\mathbb{P}(X_n \geq k) = \mathrm{O}\left(n^{-2}\right)$. Observe that with use of the upper bound for d_k (2.4), we obtain an upper bound for the probability of X_n being bigger or equal k that holds for any $k \geq 1$

$$\mathbb{P}(X_n \geq k) \leq \left(\frac{2}{n}\right)^{k-1} + \mathrm{O}\left(n^{\frac{1}{2}-k}\right). \tag{2.6}$$

Thus $\sum_{k\geq 3}\mathbb{P}(X_n \geq k) = \mathrm{O}\left(n^{-2}\right)$ and we get the result for the expected value of the protection number of uniformly chosen random PORT.

In order to compute the variance of X_n we will again utilize the fact that X_n is non-negative and integer valued random variable. This allows us to use the following well known formula for the second moment:

$$\begin{aligned}\mathbb{E}\left[X_n^2\right] &= \sum_{k\geq 1}(2k-1)\mathbb{P}(X_n \geq k) \\ &= 1 + \frac{6}{n} + \mathrm{O}\left(\frac{1}{n^{\frac{3}{2}}}\right) + \sum_{k\geq 3}(2k-1)\mathbb{P}(X_n \geq k).\end{aligned}$$

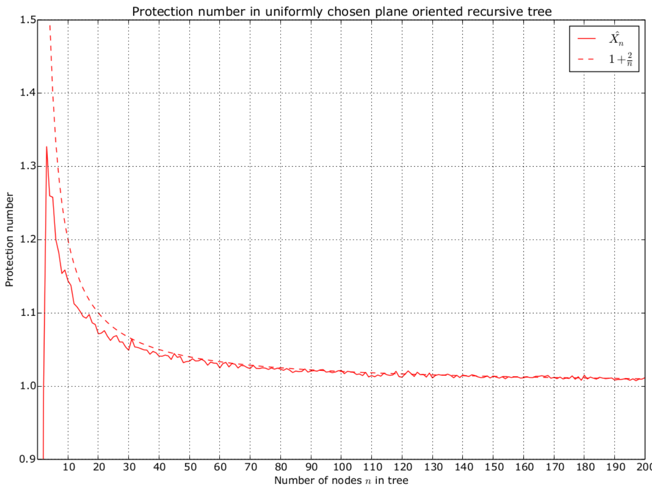

Figure 4: Plot of estimated mean of X_n and $1 + \frac{2}{n}$.

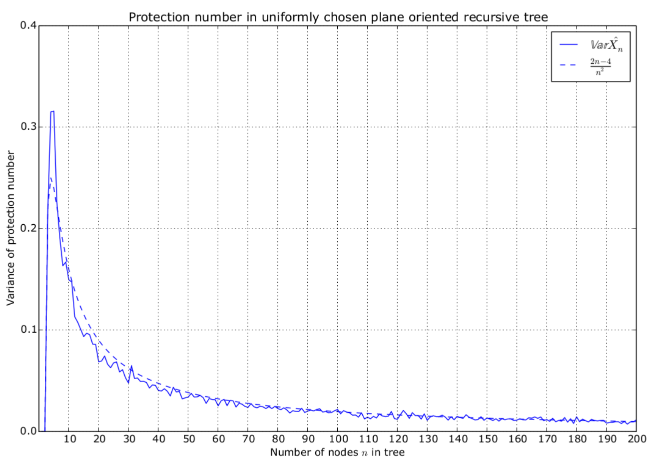

Figure 5: Plot of estimated variance of X_n and $\frac{2}{n} - \frac{4}{n^2}$.

Again using the upper bound for the probability of X_n being bigger or equal k (2.6) we have that $\sum_{k\geq 3}(2k-1)\mathbb{P}\left(X_n \geq k\right) = \mathrm{O}\left(n^{-2}\right)$ and therefore

$$\mathbb{E}\left[X_n^2\right] = 1 + \frac{6}{n} + \mathrm{O}\left(\frac{1}{n^{\frac{3}{2}}}\right).$$

Squaring $\mathbb{E}\left[X_n\right]$ and using standard formula for the variance $\mathbb{V}\left[X_n\right] = \mathbb{E}\left[X_n^2\right] - \mathbb{E}\left[X_n\right]^2$ we get the desired result.

We have checked these results experimentally. For each $n = 1, 2, \ldots, 200$ we generate 2500 random PORTs and calculate mean and variance of their protection numbers. Outcomes of these experiments are shown in Figure 4 and 5.

3 Protection Number of Non-plane Recursive Trees

Let $\mathcal{U}_n$ be the set of non-plane recursive trees with n nodes. Let a random variable Y_n denote the protection number of randomly chosen non-plane recursive tree with n nodes (we assume uniform distribution on the elements of the set $\mathcal{U}_n$).

Let $\mathcal{U}_{n,k} \subseteq \mathcal{U}_n$ be a set of all non-plane recursive trees with n nodes that are k-protected and let $u_{n,k} = |\mathcal{U}_{n,k}|$ denote its cardinality. Observe that $\mathcal{U}_n = \mathcal{U}_{n,0} \supseteq \mathcal{U}_{n,1} \supseteq \mathcal{U}_{n,2} \supseteq \ldots$. Let us introduce labelled combinatorial classes $(\mathcal{U}_{\geq k}, |\cdot|)$, where $\mathcal{U}_{\geq k} = \bigcup_{n\geq 0}\mathcal{U}_{n,k}$ is the set of all k-protected non-plane recursive trees and the size function $|\cdot|$ counts the number of nodes in a tree. Let $U_k(z)$ be an exponential generating function of the class $\mathcal{U}_{\geq k}$:

$$U_k(z) = \sum_{n\geq 0}|\mathcal{U}_{n,k}|\,\frac{z^n}{n!} = \sum_{n\geq 0}u_{n,k}\frac{z^n}{n!}.$$

We can specify a k-protected non-plane recursive tree as a pair of root's label and non-empty set of subtrees whose roots are $(k-1)$-protected. There is also an order constraint – the smallest label belongs to the root. From this specification we get the following symbolic equation:

$$\mathcal{U}_{\geq k} = \mathcal{Z}^{\Box} \star \textsc{Set}_{\geq 1}(\mathcal{U}_{\geq k-1}).$$

The difference with PORTs is that in the specification we replace "non-empty sequence" of subtrees by "non-empty set" of subtrees, which means that in case of non-plane recursive trees the order of subtrees does not count.

Notice that $\sum_{n\geq 1}(n-1)!\frac{z^n}{n!} = -\ln(1-z)$. This leads us to the following recursive system of equations

$$\begin{cases} U_k(z) = \int_0^z e^{U_{k-1}(t)}dt - z & \text{for } k \geq 1, \\ U_0(z) = -\ln(1-z). \end{cases}$$

Similarly like in case of PORTs, using standard integration techniques we are able to calculate $U_1(z) = -\ln(1-z) - z = U_0(z) - z$ and again this result is obvious when one realizes that the only tree which is 0-protected and is not 1-protected is a single node.

The generating function of the number of 2-protected non-plane recursive trees can be expressed with use of the special function

$$U_2(z) = \int_0^z \frac{dt}{e^t\cdot(1-t)} - z = \frac{\mathtt{Ei}(1) - \mathtt{Ei}(1-z)}{e} - z,$$

where $\mathtt{Ei}(z)$ is the exponential integral function defined for real non-zero values z as $\mathtt{Ei}(z) = -\int_{-z}^{\infty}\frac{e^{-t}}{t}dt$ and for complex values z such that $|arg(z)| < \pi$ as

$\mathtt{Ei}(z) = -\mathtt{E}_1(-z) \pm i\pi$, where $\mathtt{E}_1(z) = \int_z^\infty \frac{e^{-t}}{t} dt$ (for more see [1, Ch. 5]). Its value in 1 can be numerically obtained $\mathrm{Ei}(1) = 1.89511781\ldots$ (see OEIS A091725). Moreover, for complex values z such that $|arg(z)| < \pi$ the exponential integral $E_1(z)$ admits asymptotic expansion $E_1(z) = -\gamma - \ln(z) - \sum_{n\geq 1} \frac{(-1)^n z^n}{n \cdot n!}$.

From the above we can obtain first coefficients of $U_2(z)$ function: $0, 0, 0, 1, 2, 9, 44, 265, 1854, 14833, 133496, \ldots$. For $n \geq 2$ it is the sequence OEIS A000166 i.e. the number of permutations of $n-1$ elements with no fixed points or so called derangements. This connection is due to the bijection between non-plane recursive trees of size n and permutations of $n-1$ elements. This bijection transfers some shape characteristics. For example, the root degree of the non-plane recursive tree corresponds to the number of cycles in the cycle decomposition and the subtree sizes of the root correspond to the sizes of the cycles (for more about the bijection see [19, Ch. 6]). So, we see that the 2-protected non-plane recursive trees of size n are trees with all subtrees of the root of size at least 2. But those corresponds to the permutations of $n-1$ elements with all cycles of length at least 2 in the cycle decomposition, thus permutations with no fixed points. Number of derangements of m elements is well studied and denoted by $!m = \left\lfloor \frac{m!}{e} + \frac{1}{2} \right\rfloor$. So, we immediately obtain the number of 2-protected non-plane recursive trees of size n as

$$u_{n,2} = \begin{cases} 0 & \text{if } n \leq 2, \\ \left\lfloor \frac{(n-1)!}{e} + \frac{1}{2} \right\rfloor = \frac{(n-1)!}{e}(1 + \mathrm{o}(1)) & \text{oth.} \end{cases}$$

Unfortunately it seems that such a nice connection does not carry over even to the number of 3-protected non-plane recursive trees.

Now, let us state few results concerning Y_n the protection number of randomly chosen non-plane recursive tree with n nodes.

Theorem 3.1. *For $k \geq 1$, probability that uniformly chosen non-plane recursive tree with n nodes is k-protected equals*

$$\mathbb{P}(Y_n \geq k) = \begin{cases} 1 & \text{if } k = 1, \\ \frac{1}{e} + \mathrm{o}(1) & \text{if } k = 2, \\ \mathrm{O}\left(n^{2+\frac{1}{e}-k}\right) & \text{if } k \geq 3, \end{cases} \quad \text{as } n \to \infty.$$

Observe that having uniform distribution among the trees of size n, from the previous observations we easily get the result for $k = 1$ and 2. Moreover, we have that

$$\mathbb{P}(Y_n \geq k) = [z^n]\; zU_k'(z).$$

So, for $k \geq 3$ a similar analysis like in case of plane oriented recursive trees applies. Observe that neither

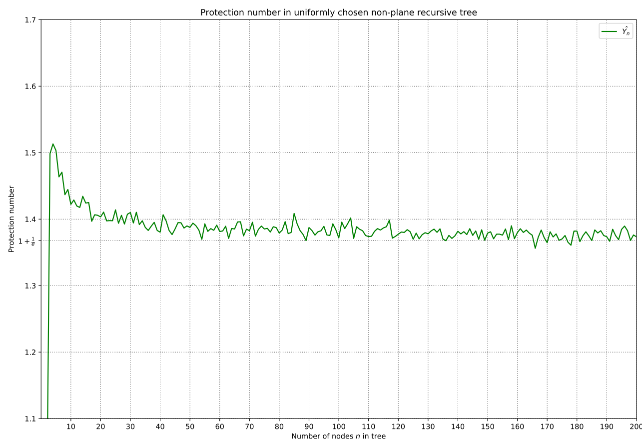

Figure 6: Plot of estimated mean of Y_n.

integration operator nor $e^{f(z)}$ operator influences the dominant singularity position thus it stays at $\rho = 1$. Those operators influences only the nature of the dominant singularity. Rest of the proof follows so called *Exp-log schema* (see [20, Thm. VII.1]) and again *Singular Integration Theorem* (see [20, Thm. VI.9]).

In this extended abstract we will skip the details of the analysis in non-plane case and move to the consequences of the above theorem.

Corollary 3.1. *Mean and variance of the protection number of uniformly chosen random non-plane recursive tree equals*

$$\mathbb{E}[Y_n] = 1 + \frac{1}{e} + \mathrm{O}\left(\frac{1}{n^{1-\frac{1}{e}}}\right),$$
$$\mathbb{V}[Y_n] = \frac{1}{e} - \frac{1}{e^2} + \mathrm{O}\left(\frac{1}{n^{1-\frac{1}{e}}}\right).$$

Again, this corollary follows the fact that Y_n is non-negative and integer random variable for which we can use the well know formulas for the first and second moment.

We have checked these results experimentally. For each $n = 1, 2, \ldots, 200$ we generate 2500 random non-plane recursive trees and calculate mean and variance of their protection numbers. Outcomes of these experiments are shown in Figure 6 and 7.

4 Conclusions

In this paper we have studied the protection number of plane oriented recursive trees and non-plane recursive trees. We have derived the probability distribution of the considered parameter as well as its mean and variance. In Table 1 we can see comparison of the mean of the protection number in a different models of random

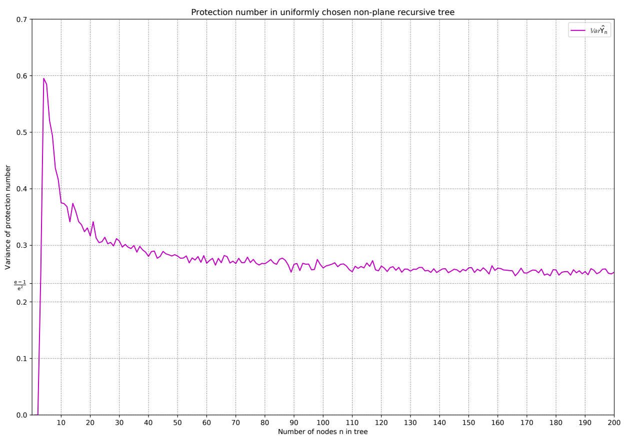

Figure 7: Plot of estimated variance of Y_n.

trees. Observe that the mean of protection number of non-plane recursive trees lies almost in the middle between plane oriented recursive trees and unlabelled plane trees.

The next thing that we would like to consider is the protection number of a randomly chosen node in a random recursive tree. Experiments reveal that in case of both considered models the mean value of this parameter is about $0.6 - 0.7$. In [30] Heuberger and Prodinger were able to calculate probability distribution of the protection number of randomly chosen node in case of unlabelled plane trees. They use simple but convenient bijection between a tree t and a k-protected vertex v in this tree and a k-protected tree t_1 whose root is merged with a leaf of another tree t_2. It seems that such simple bijection does not work in case of recursive trees because one has to take into consideration the order constraints on the labels of the vertices. So, this task seems to be more challenging in recursive trees case.

There are also other interesting parameters concerning protection number of a vertex. For instance, experiments reveal that the maximal protection number of a node in PORTs and non-plane recursive trees are about $1.5 \log\log n$ and $2 \log\log n$ respectively.

Since the two models we have considered here fall into the category of linear preferential attachment trees introduced by Pittel in [10], the next step of studies could also be to find the protection number for this broader class of trees. In the evolution process of the linear preferential attachment tree for each new node, its parent is chosen among the existing nodes with the probability of choosing a node v proportional to $w_{d(v)}$, where $d(v)$ is the out-degree of v and $w_k = \chi \cdot k + \rho$ for some real parameters χ and $\rho > 0$. Thus the random plane oriented recursive tree is obtained for $\chi = \rho = 1$ and the random non-plane recursive tree is obtained for

Tree model	Mean of the protection number	Reference
Unlabelled plane trees	$1.62297 + \frac{0.13118}{n} + \mathrm{O}\left(\frac{1}{n^{\frac{3}{2}}}\right)$	[29, 30]
Plane oriented recursive trees	$1 + \frac{2}{n} + \mathrm{O}\left(\frac{1}{n^{\frac{3}{2}}}\right)$	This paper
Non-plane recursive trees	$1 + \frac{1}{e} + \mathrm{O}\left(\frac{1}{n^{1-\frac{1}{e}}}\right)$	This paper
PATRICIA trees	$\sim \log_2 n$	[7]
d-ary recursive trees	$\sim \alpha_d \log n$	[12, 19]

Table 1: Comparison of the mean of protection number in a different tree models.

$\chi = 0$ and $\rho = 1$. So, the desired results would be the probability distribution as well as mean and variance of the protection number of linear preferential attachment tree that are some functions of χ and ρ.

References

[1] Abramowitz, M. and Stegun, I., *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York: Dover. ISBN 0-486-61272-4, 1964.

[2] Dondajewski, M. and Szymański, J., *On the distribution of vertex-degrees in a strata of a random recursive tree*. In Bull. Acad. Polon. Sci. Ser. Sci. Math., 30(5-6):205-209, 1982.

[3] Szymański, J., *On a nonuniform random recursive tree*. In Annals of Discrete Math. 33:297-306, 1987.

[4] Szymański, J., *On the maximum degree and the height of a random recursive tree*. In Random graphs '87 (Poznań, 1987), pages 313-324. Wiley, Chichester, 1990.

[5] Aldous, D., *Asymptotic fringe distributions for general families of random trees*. In Ann. Appl. Probab. 1(2):228-266, 1991.

[6] Bergeron, F., Flajolet, P. and Salvy, B., *Varieties of increasing trees*. In CAAP '92, Lecture Notes in Computer Science, vol 581. Springer, Berlin, Heidelberg.

[7] Devroye, L., *A note on the probabilistic analysis of patricia trees*. In Random Struct. Alg., 3:203-214, 1992.

[8] Mahmoud, H. M., *Distances in random plane-oriented recursive trees*. In Journal of Computational and Applied Mathematics 41:237-245, 1992.

[9] Smythe, R. T. and Mahmoud, H. M., *A survey of recursive trees*. In Theory of Probability and Mathematical Statistics, 51:1-27, 1994.

[10] Pittel, B., *Note on the heights of random recursive trees and random m-ary search trees*. In Random Structures Algorithms, 5(2):337-347, 1994.

[11] Prodinger, H., *Depth and path length of heap ordered trees*. In International Journal of Foundations of Computer Science 7:293-299, 1996.

[12] Drmota. M., *The saturation level in binary search tree*. In Mathematics and computer science (Versailles, 2000), Trends Math., pages 41-51, 2000.

[13] Fuchs, M., Hwang, H.-K. and Neininger, R., *Profiles of random trees: limit theorems for random recursive trees and binary search trees*. In Algorithmica, 46(3-4):367-407, 2006.

[14] Kuba, M. and Panholzer, A., *On the degree distribution of the nodes in increasing trees*. In Journal of Combinatorial Theory, Series A, 114(4):597-618, 2007.

[15] Hwang, H.-K., *Profiles of random trees: plane-oriented recursive trees*. In Random Structures Algorithms, 30(3):380-413, 2007.

[16] Broutin, N., Devroye, L., McLeish, E. and de la Salle, M., *The height of increasing trees*. In Random Structures Algorithms, 32(4):494-518, 2008.

[17] Cheon, G.-S. and Shapiro, L., *Protected points in ordered trees*. In Appl. Math. Lett., 21(5):516-520, 2008.

[18] Drmota, M., *The height of increasing trees*. In Annals of Combinatorics, 12(4):373-402, 2009.

[19] Drmota, M., *Random trees: an interplay between combinatorics and probability*. Springer Science & Business Media, 2009.

[20] Flajolet, P. and Sedgewick, R., *Analytic Combinatorics*. Cambridge University Press, 2009.

[21] Janson, S., Kuba, M. and Panholzer, A., *Generalized Stirling permutations, families of increasing trees and urn models*. In Journal of Combinatorial Theory, Series A, 118(1):94-114, 2011.

[22] Mansour, T., *Protected points in k-ary trees*. In Appl. Math. Lett., 24(4):478-480, 2011.

[23] Du, R. R. X. and Prodinger, H., *Notes on protected nodes in digital search trees*. In Appl. Math. Lett. 25:1025âĂŞ1028, 2012.

[24] Holmgren, C. and Janson, S., *Limit laws for functions of fringe trees for binary search trees and recursive trees*. In Electronic J. Probability 20(4):1-51, 2015.

[25] Mahmoud, H. M. and Ward, M. D., *Asymptotic distribution of two-protected nodes in random binary search trees*. In Appl. Math. Lett. 25:2218-2222, 2012.

[26] Devroye, L. and Janson, S., *Protected nodes and fringe subtrees in some random trees*. In Electronic Communications in Probability 19(6):1-10, 2014.

[27] Mahmoud, H. M. and Ward, M. D., *Asymptotic properties of protected nodes in random recursive trees*. In J. Appl. Probab. 52(1):290-297, 2015.

[28] Bóna, M. and Pittel, B., *On a random search tree: Asymptotic enumeration of vertices by distance from leaves*. In Advances in Applied Probability, 49(3):850-876, 2017.

[29] Copenhaver, K. *k-Protected Vertices in Unlabeled Rooted Plane Trees*. In Graphs and Combinatorics 33(2):347-355, 2017.

[30] Heuberger C., and Prodinger, H., *Protection number in plane trees*. In Applicable Analysis and Discrete Mathematics Vol. 11:314-326, 2017.