On the average number of nodes covering a given number of leaves in an unordered binary tree

Nozomu Ochiumi[1*] Fumiaki Kanazawa[2†‡]
Masahiro Yanagida[1§] Yasuichi Horibe[1¶]

[1] *Department of Mathematical Information Science, Faculty of Science, Tokyo University of Science, 1-3 Kagurazaka, Shinjuku-ku, Tokyo 162-8601, Japan*

[2] *Japan Patent Office, 3-4-3 Kasumigaseki, Chiyoda-ku, Tokyo 100-8915, Japan*

Abstract

The covering number for a subset of leaves in a finite rooted tree is defined as the number of subtrees which remain after deleting all the paths connecting the root and the other leaves. We find the formula for the total sum (hence the average) of the covering numbers for a given subset of labeled leaves over all unordered binary trees with n leaves.

Keyword: covering number, unordered binary tree, recursion, generating function method

1 Introduction

Let J be a subset of leaves in a finite rooted tree T with leaf-set U. If we delete all the paths (and all the edges incident to them) that connect the root and the leaves in $U \setminus J$, then there remains a forest comprising, say,

*j1105702@ed.kagu.tus.ac.jp (Corresponding author)
†kfumiaki@ieee.org
‡This work was done while the author was in the University of Tsukuba.
§yanagida@rs.kagu.tus.ac.jp
¶horibe@rs.kagu.tus.ac.jp

c subtrees of T. We may say that, in the whole tree T, the c nodes, the roots of these subtrees, *cover* or dominate J (and only J), and call c the *covering number* for J.

The "cover" concept for rooted trees seems to be originated in the works [1][6] on a certain cryptographic key-management problem of a broadcast type with $|U|$ users, and the covering number can be considered a new combinatorial topic in the theory of rooted trees. We believe that it is worthwhile to investigate covering numbers from the view point of combinatorics and derive mathematical results such as the distribution and expected value of the covering number for J with respect to a certain probability measure $P(T)$, $T \in \mathcal{T}$, where $\mathcal{T}$ is the set of all binary trees (either ordered or unordered) with n $(= |U|)$ leaves. The main purpose of this paper is to find an explicit formula for the average covering number for J in the case where T is an unordered binary tree with labeled leaves and $P(T)$ is uniform, i.e., $P(T) = |\mathcal{T}|^{-1}$ (see [2] for a corresponding study on the completely balanced binary tree with 2^k leaves).

As is described in [5], an unordered binary tree with n labeled leaves $1, 2, \ldots, n$ is a graphic representation of a "binary total partition" of $U = \{1, 2, \ldots, n\}$; partition U (the root) into two non-empty subsets (unordered two children of the root), similarly bipartition each of these subsets, ..., continued until we have n singleton sets (n leaves). Denote by $\mathcal{T}_U$ the set of all such binary trees having n leaves and put $b_n = |\mathcal{T}_U|$, then it is shown that $b_1 = 1$ and

$$b_n = \frac{1}{2}\sum_{k=1}^{n-1}\binom{n}{k} b_k b_{n-k}, \quad n \geqslant 2,$$

which leads us to the formula $b_n = (2n-3)!!$ (as was originally given in [3]), where $n!!$ means the double-factorial (define $(-1)!! = 1$ and $b_0 = 0$).

Let $c_T(J)$ be the covering number for a subset $J \subset U$ of leaves in a finite rooted tree $T \in \mathcal{T}_U$ and define $c_T(\emptyset) = 0$. We are interested in finding the average covering number for J of size k, defined as $\frac{a_{n,k}}{b_n}$, where

$$a_{n,k} = \sum_{T \in \mathcal{T}_U} c_T(J), \quad 0 \leqslant k \leqslant n = |U|, \quad J \in \binom{U}{k}.$$

Note that the average covering number is independent of the choice of J, that is,

$$\sum_{T \in \mathcal{T}_U} c_T(J) = \sum_{T \in \mathcal{T}_U} c_T(J'), \quad J, J' \in \binom{U}{k}.$$

In Section 2 we derive a recurrence relation for $a_{n,k}$ and give an explicit expression for the average covering number. In Section 3, asymptotic behavior of $\frac{a_{n,k}}{b_n}$ is shown by fixing $\frac{k}{n}$ and taking n large.

2 A recursion and the general term

Let $a_{n,k} = 0$ $(k > n)$ for convenience sake. It is clear that $a_{n,n} = b_n$ $(n \geqslant 1)$ by the definition of $a_{n,k}$. We first show that $a_{n,k}$'s satisfy the following recursion.

Theorem 1. *For* $1 \leqslant k \leqslant n-1$,

$$a_{n,k} = \sum_{l=1}^{n-1} \sum_{i=1}^{k} \binom{k}{i} \binom{n-k}{l-i} (2(n-l)-3)!! a_{l,i}$$
$$= \sum_{i=1}^{k} \sum_{j=0}^{n-k} \binom{k}{i} \binom{n-k}{j} (2(n-i-j)-3)!! a_{i+j,i}.$$

Proof. Let $|U| = n$, and T_1, T_2 be the two subtrees of the root of $T \in \mathcal{T}_U$. Define V and its complement V^c as the sets of leaves of T_1 and T_2 respectively. Then $c_T(J) = c_{T_1}(J \cap V) + c_{T_2}(J \cap V^c)$ holds for $J \in \binom{U}{k}$ $(1 \leqslant k \leqslant n-1)$, and we have

$$a_{n,k} = \sum_{T \in \mathcal{T}_U} c_T(J)$$
$$= \frac{1}{2} \sum_{l=1}^{n-1} \sum_{V \in \binom{U}{l}} \sum_{T_1 \in \mathcal{T}_V} \sum_{T_2 \in \mathcal{T}_{V^c}} (c_{T_1}(J \cap V) + c_{T_2}(J \cap V^c))$$
$$= \frac{1}{2} \sum_{l=1}^{n-1} \sum_{V \in \binom{U}{l}} \left(|\mathcal{T}_{V^c}| \sum_{T_1 \in \mathcal{T}_V} c_{T_1}(J \cap V) + |\mathcal{T}_V| \sum_{T_2 \in \mathcal{T}_{V^c}} c_{T_2}(J \cap V^c) \right)$$
$$= \frac{1}{2} \sum_{l=1}^{n-1} \sum_{V \in \binom{U}{l}} \left(a_{l,|J\cap V|} b_{n-l} + a_{n-l,|J \cap V^c|} b_l \right).$$

Noting that $\left|\{V \in \binom{U}{l} \mid |J \cap V| = i\}\right| = \binom{k}{i}\binom{n-k}{l-i}$ and $\left|\{ V \in \binom{U}{l} \mid |J \cap V^c|\right.$

$= k - i \}| = \binom{k}{k-i}\binom{n-k}{n-l-k+i}$ for $0 \leqslant i \leqslant k$, we obtain

$$\begin{aligned} a_{n,k} &= \frac{1}{2}\sum_{l=1}^{n-1}\sum_{i=0}^{k}\left(\binom{k}{i}\binom{n-k}{l-i}a_{l,i}b_{n-l}\right. \\ &\qquad\qquad \left. + \binom{k}{k-i}\binom{n-k}{n-k-l+i}a_{n-l,k-i}b_l\right) \\ &= \sum_{l=1}^{n-1}\sum_{i=1}^{k}\binom{k}{i}\binom{n-k}{l-i}a_{l,i}b_{n-l} \qquad (1) \\ &= \sum_{i=1}^{k}\sum_{j=0}^{n-k}\binom{k}{i}\binom{n-k}{j}a_{i+j,i}b_{n-i-j}. \end{aligned}$$

□

Next we derive the formula for the general term $a_{n,k}$ by the generating function method.

Theorem 2. *For $n \geqslant 2$,*

$$a_{n,k} = (2(n-k)-1)!!\left(\frac{(2n-2)!!}{(2(n-k)-2)!!} - \frac{(2n-3)!!}{(2(n-k)-3)!!}\right),\ 1 \leqslant k \leqslant n-1.$$

We remark that $a_{n,n-1} = (2n-2)!! - (2n-3)!!$. This special case is mentioned in [4, A129890] without its sources.

Proof. Let $A_k(x) = \sum_{n\geqslant k}\frac{a_{n,k}}{(n-k)!}x^{n-k}$ for $k \geqslant 1$ and $B(x) = \sum_{n\geqslant 0}\frac{b_n}{n!}x^n$ $(= 1-\sqrt{1-2x})$. Then the s-th derivative of $B(x)$ is $B^{(s)}(x) = \sum_{n\geqslant s}\frac{b_n}{(n-s)!}x^{n-s}$. The coefficient of x^{n-k} in $A_i(x)B^{(k-i)}(x)$ is $\sum_{l=i}^{n-k+i}\frac{a_{l,i}}{(l-i)!}\frac{b_{n-l}}{(n-l-(k-i))!}$, and

$$\begin{aligned} \frac{a_{n,k}}{(n-k)!} &= \sum_{i=1}^{k}\sum_{l=i}^{n-k+i}\frac{\binom{k}{i}\binom{n-k}{l-i}}{(n-k)!}a_{l,i}b_{n-l} \\ &= \sum_{i=1}^{k}\binom{k}{i}\sum_{l=i}^{n-k+i}\frac{a_{l,i}}{(l-i)!}\frac{b_{n-l}}{(n-l-(k-i))!}, \end{aligned}$$

for $n \geqslant k+1$ by (1), hence we have

$$\sum_{n\geqslant k+1}\frac{a_{n,k}}{(n-k)!}x^{n-k} = \sum_{n\geqslant k+1}\sum_{i=1}^{k}\binom{k}{i}\sum_{l=i}^{n-k+i}\frac{a_{l,i}}{(l-i)!}\frac{b_{n-l}}{(n-l-(k-i))!}x^{n-k},$$

so that

$$A_k(x) - a_{k,k} = \sum_{i=1}^{k}\binom{k}{i}\left(A_i(x)B^{(k-i)}(x) - a_{i,i}b_{k-i}\right),\ k \geqslant 1.$$

Since $a_{i,i} = b_i$ and $\sum_{i=1}^{k} \binom{k}{i} a_{i,i} b_{k-i} = \sum_{i=1}^{k-1} \binom{k}{i} b_i b_{k-i} = 2b_k$ $(k \geqslant 2)$, we have

$$\frac{A_k(x)}{k!} = \sum_{i=1}^{k} \frac{A_i(x)}{i!} \frac{B^{(k-i)}(x)}{(k-i)!} - \frac{b_k}{k!}, \ k \geqslant 2.$$

Let $A(x,y) = \sum_{k\geqslant 1} \frac{A_k(x)}{k!} y^k = \sum_{k\geqslant 1} \sum_{n\geqslant k} \frac{a_{n,k}}{(n-k)!k!} x^{n-k} y^k$ and $B(x,y) = \sum_{k\geqslant 0} \frac{B^{(k)}(x)}{k!} y^k$ $(= B(x+y))$. Then the coefficient of y^k in $A(x,y)B(x,y)$ is $\sum_{i=1}^{k} \frac{A_i(x)}{i!} \frac{B^{(k-i)}(x)}{(k-i)!}$, hence we have

$$\begin{aligned} A(x,y) - A_1(x)y &= \sum_{k\geqslant 2} \frac{A_k(x)}{k!} y^k \\ &= \sum_{k\geqslant 2} \left(\sum_{i=1}^{k} \frac{A_i(x)}{i!} \frac{B^{(k-i)}(x)}{(k-i)!} - \frac{b_k}{k!} \right) y^k \\ &= A(x,y)B(x,y) - A_1(x)B(x)y - B(y) + y. \end{aligned}$$

Since $a_{n,1} = b_n$ and $A_1(x) = \sum_{n\geqslant 1} \frac{b_n}{(n-1)!} x^{n-1} = B'(x) = \frac{1}{\sqrt{1-2x}}$, we have

$$\begin{aligned} A(x,y) &= \frac{(1-B(x))B'(x)y - B(y) + y}{1 - B(x+y)} \\ &= \left((1-2y)^{\frac{1}{2}} - (1-2y)\right)(1-2y-2x)^{-\frac{1}{2}}. \end{aligned}$$

The coefficient of $x^t y^k$ in $A(x,y)$ is

$$\frac{(2t-1)!!}{t!k!} \left(\frac{(2t+2k-2)!!}{(2t-2)!!} - \frac{(2t+2k-3)!!}{(2t-3)!!} \right) \tag{2}$$

because we have

$$\begin{aligned} (1-2y-2x)^{-\frac{1}{2}} &= (1-2y)^{-\frac{1}{2}} \sum_{t\geqslant 0} \binom{-\frac{1}{2}}{t} \left(-\frac{2x}{1-2y} \right)^t \\ &= \sum_{t\geqslant 0} \frac{(2t-1)!!}{t!} x^t (1-2y)^{-t-\frac{1}{2}} \end{aligned}$$

and

$$
\begin{aligned}
&(1-2y)^{-t}-(1-2y)^{-t+\frac{1}{2}} \\
&=\sum_{k\geqslant 0}\left(\binom{-t}{k}-\binom{-t+\frac{1}{2}}{k}\right)(-2y)^k \\
&=\sum_{k\geqslant 1}\Bigg(t(t+1)\cdots(t+k-1) \\
&\qquad -\left(t-\frac{1}{2}\right)\left(t+\frac{1}{2}\right)\left(t+\frac{3}{2}\right)\cdots\left(t-\frac{1}{2}+k-1\right)\Bigg)\frac{(2y)^k}{k!} \\
&=\sum_{k\geqslant 1}\left(\frac{(2t+2k-2)!!}{(2t-2)!!}-\frac{(2t+2k-3)!!}{(2t-3)!!}\right)\frac{y^k}{k!}.
\end{aligned}
$$

Substituting t with $n-k$ in (2) finishes the proof. □

3 Asymptotic curve

The ratio of the number of nodes necessary to cover $J \in \binom{U}{k}$ to the number $2n-1$ of all nodes is $\frac{c_T(J)}{2n-1}$ for tree $T \in \mathcal{T}_U$, and its average over $\mathcal{T}_U$ is $\frac{a_{n,k}}{(2n-1)b_n}$. We fix the ratio $\frac{k}{n}$ between $|J|$ and $|U|$ to ρ and consider the limit of the average $\frac{c_T(J)}{2n-1}$ as $n \to \infty$. By Theorem2

$$
\begin{aligned}
\frac{a_{n,k}}{b_n} &= \frac{(2(n-k)-1)!!}{(2n-3)!!}\left(\frac{(2n-2)!!}{(2(n-k)-2)!!}-\frac{(2n-3)!!}{(2(n-k)-3)!!}\right) \\
&= (2(n-k)-1)\left(\frac{(2(n-k)-3)!!\,(2n-2)!!}{(2(n-k)-2)!!\,(2n-3)!!}-1\right).
\end{aligned}
$$

By Stirling's formula, we have

$$
\frac{(2t-1)!!}{(2t)!!}=\frac{(2t)!}{2^{2t}(t!)^2}\sim\frac{2\sqrt{\pi t}\left(\frac{2t}{e}\right)^{2t}}{2^{2t}2\pi t\left(\frac{t}{e}\right)^{2t}}=\frac{1}{\sqrt{\pi t}}.
$$

Therefore, we obtain

$$
\begin{aligned}
\frac{a_{n,k}}{(2n-1)b_n} &\sim \frac{2(n-k)-1}{2n-1}\left(\sqrt{\frac{n-1}{n-k-1}}-1\right) \\
&\to (1-\rho)\left(\frac{1}{\sqrt{1-\rho}}-1\right) \\
&= \sqrt{1-\rho}\left(1-\sqrt{1-\rho}\right).
\end{aligned}
$$

Consequently, the line through the points $\left(\frac{k}{n}, \frac{a_{n,k}}{(2n-1)b_n}\right)$ $(k = 0, 1, \ldots, n)$ approaches the curve $f(\rho) = \sqrt{1-\rho}(1-\sqrt{1-\rho})$ which attains the maximal value $\frac{1}{4}$ when $\rho = \frac{3}{4}$.

References

[1] D. Naor, M. Naor, J. Lotspiech, Revocation and tracing schemes for stateless receivers, Advances in Cryptology—CRYPTO 2001 (Santa Barbara, CA), vol. 2139 of Lecture Notes in Comput. Sci., Springer, Berlin, 2001, pp. 41–62.

[2] E. C. Park, I. F. Blake, On the mean number of encryptions for tree-based broadcast encryption schemes, J. Discrete Algorithms, 4 (2) (2006) 215–238.

[3] E. Schröder, Vier kombinatorische Probleme, Z. für Math. Phys., 15 (1870) 361–376.

[4] N. J. A. Sloane, The On-Line Encyclopedia of Integer Sequences, http://www.research.att.com/~njas/sequences

[5] R. P. Stanley, Enumerative Combinatorics. Vol. 2, Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge, 1999.

[6] C. K. Wong, M. G. Gouda, S. S. Lam, Secure group communications using key graphs, IEEE/ACM Trans. Networking, 8 (1) (2000) 16–30.