A Note on the Weight of Multicast Shortest Path Trees

Piet Van Mieghem*

July 26, 2007

Abstract

The weight of a multicast shortest path tree in the complete graph with exponential link weights is expressed as a random variable. The new framework elegantly allows the computation of the average multicast weight.

1 Introduction

We consider the problem of computing the cost of multicasting information from a source to m different nodes in a network containing $N+1$ nodes. We assume that the multicast tree rooted at the source, which we label by 0, is the shortest path tree to the m different nodes. In addition, these m multicast member nodes are supposed to be uniformly distributed among the N nodes (different from the source) in the network. Both assumptions are realistic and have been justified earlier [2].

We further confine to a complete graph K_{N+1} with exponentially distributed link weights with mean 1. The present note is an extension of [3] where the weight $W_N(m)$ to all $m = N$ possible multicast members is shown to tend to a Gaussian,

$$\sqrt{N}\left(W_N\left(N\right) - \zeta\left(2\right)\right) \xrightarrow{d} N\left(0, \sigma^2_{\text{SPT}}\right),$$

where $\sigma^2_{\text{SPT}} = 4\zeta\left(3\right)$. In addition, we recompute the average weight of the multicast tree $\mathbb{E}[W_N(m)]$ (see eq. (10)) more elegantly than in our previous work [1]. Beside the novel mathematical framework, the work is motivated by the new peer-to-peer protocol HMTP that attaches users to the tree in order to minimize the total cost $W_N\left(m\right)$. Moreover, in IPTV, an operator is businesswise interested in the cost difference $W_N\left(m+k\right) - W_N\left(m\right)$ by adding k additional customers to a TV channel, given a multicast tree of that TV channel with m multicast members. The number of hops to the nearest of m peers is analyzed in [4].

2 The weight of the SPT to m nodes

The shortest path tree in the complete graph with exponential link weight is a uniform recursive tree. A uniform recursive tree (URT) of size N is a random tree rooted at node 0 where at each stage of the growth process a new node is attached uniformly to one of the existing nodes until the total number

*P.VanMieghem@ewi.tudelft.nl, Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, P.O. Box 5031, 2600 GA Delft, The Netherlands.

of nodes is equal to N. The nice properties of the shortest path search process in the complete graph with exponential link weights are explained in [5, Chapter 16]. The path search or discovery process can be described as a Markov process. From the Markov discovery process , the discovery time to the k^{th} discovered node from the root equals

$$v_k = \sum_{n=1}^{k} \tau_n, \tag{1}$$

where the inter-attachment times $\tau_1, \tau_2, \cdots, \tau_k$ are independent, exponentially distributed random variables with parameter $\lambda_n = n(N+1-n)$ for $1 \le n \le k$. An arbitrary uniform recursive tree consisting of $N+1$ nodes and with the root labeled by zero can be represented as

$$(0 \longleftarrow 1)\,(N_2 \longleftarrow 2)\ldots(N_N \longleftarrow N)$$

where $(N_j \longleftarrow j)$ means that the j^{th} discovered node is attached to node $N_j \in \{0, \ldots, j-1\}$. Hence, N_j is the predecessor of j and this relation is indicated by $\longleftarrow$. The weight $W_N(m)$ of the SPT from the root 0 to m other nodes is with (1) and $v_0 = 0$ and $N_1 = 0$,

$$W_N(m) = \sum_{j=1}^{N} (v_j - v_{N_j}) \mathbf{1}_{j \in T_m} = \sum_{j=1}^{N} \mathbf{1}_{j \in T_m} \sum_{n=N_j+1}^{j} \tau_n,$$

where T_m is the subtree of the complete SPT to the m uniform nodes. In the URT, the integers N_j for $1 \le j \le N$, are independent and uniformly distributed over the interval $\{0, \ldots, j-1\}$. Following the analysis in [3], it is more convenient to use a discrete uniform random variable on $\{1, \ldots, j\}$ which we define as $A_j = N_j + 1$ such that

$$W_N(m) = \sum_{j=1}^{N} \mathbf{1}_{j \in T_m} \sum_{n=A_j}^{j} \tau_n = \sum_{j=1}^{N} \mathbf{1}_{j \in T_m} \sum_{n=1}^{j} \mathbf{1}_{\{A_j \le n\}} \tau_n$$

The set $\{A_j\}_{1 \le j \le N}$ are independent random variables with $\mathbb{P}[A_j = k] = \frac{1}{j}$ for $k \in \{1, 2, \ldots, j\}$. If we denote the m uniformly chosen multicast member nodes by $U_1, \ldots, U_m$, then node j is an element of T_m precisely when there exists an $i = 1, \ldots, m$ such that $U_i \in \mathcal{T}_j^{(N)}$, where $\mathcal{T}_j^{(N)}$ is the subtree of the complete uniform recursive tree rooted at j. We write $\vec{U}_m \cap \mathcal{T}_j^{(N)} \neq \varnothing$ for the event that there exists an $i = 1, \ldots, m$ such that $U_i \in \mathcal{T}_j^{(N)}$. Then, we have

$$W_N(m) = \sum_{j=1}^{N} \mathbf{1}_{\{\vec{U}_m \cap \mathcal{T}_j^{(N)} \neq \varnothing\}} \sum_{n=1}^{j} \mathbf{1}_{\{A_j \le n\}} \tau_n = \sum_{n=1}^{N} \tau_n \left(\sum_{j=n}^{N} \mathbf{1}_{\{\vec{U}_m \cap \mathcal{T}_j^{(N)} \neq \varnothing\}} \mathbf{1}_{\{A_j \le n\}} \right).$$

The indicator $\mathbf{1}_{\{\vec{U}_m \cap \mathcal{T}_j^{(N)} \neq \varnothing\}}$ depends on the random variables $U_1, \ldots, U_m$, which are independent of $\{A_j\}_{1 \le j \le N}$ and of the tree $\mathcal{T}_j^{(N)}$. Since the nodes $1, 2, \ldots, j$ are attached before the tree $\mathcal{T}_j^{(N)}$ starts growing at node j, the indicator $\mathbf{1}_{\{\vec{U}_m \cap \mathcal{T}_j^{(N)} \neq \varnothing\}}$ is independent of $A_1, \ldots, A_j$, but not of $A_{j+1}, \ldots, A_N$. Of course, the entire recursive tree is independent of the sequence $\tau_1, \tau_2, \ldots, \tau_N$.

To simplify the notation, we define $Z_j(m)$ as the number of elements that the vector $\vec{U_m}$ has in common with the random set $\mathcal{T}_j^{(N)}$,

$$Z_j(m) = |\vec{U}_m \cap \mathcal{T}_j^{(N)}|, \quad j = 1, 2, \ldots, N, \tag{2}$$

where we use the notation $|G|$ for the number of elements of the set G. We define for $n \in \{1, \ldots, N\}$ the random variables

$$B_n(m) = \sum_{j=n}^{N} \mathbf{1}_{\{\vec{U}_m \cap \mathcal{T}_j^{(N)} \neq \varnothing\}} \mathbf{1}_{\{A_j \leq n\}} = \sum_{j=n}^{N} \mathbf{1}_{\{Z_j(m) > 0\}} \mathbf{1}_{\{A_j \leq n\}}, \tag{3}$$

to obtain

$$W_N(m) = \sum_{n=1}^{N} B_n(m) \tau_n. \tag{4}$$

The n random variables $B_1(m), B_2(m), \ldots, B_n(m)$ are dependent, but independent from the inter-attachment times $\{\tau_j\}_{1 \leq j \leq N}$. Hence, the average weight of the multicast shortest path tree is

$$\mathbb{E}[W_N(m)] = \sum_{n=1}^{N} \frac{\mathbb{E}[B_n(m)]}{n(N+1-n)}. \tag{5}$$

We can also express the variance (and any other moment)

$$\operatorname{Var}[W_N(m)] = \sum_{n=1}^{N} \frac{(\mathbb{E}[B_n(m)])^2}{(n(N+1-n))^2} + 2 \sum_{n=1}^{N} \sum_{l=n}^{N} \frac{\operatorname{Cov}[B_n(m), B_l(m)]}{n(N+1-n)\, l(N+1-l)} \tag{6}$$

but, we did not succeed so far in computing $\operatorname{Cov}[B_n(m), B_l(m)]$.

3 The average weight $\mathbb{E}[W_N(m)]$

The average weight $\mathbb{E}[W_N(m)]$ requires the computation of $\mathbb{E}[B_n(m)]$ in which the size of the subtree $\mathcal{T}_j^{(N)}$ rooted at an arbitrary node j plays an important role. We first note that

$$\begin{aligned}
\mathbb{P}[Z_j(m) > 0] &= 1 - \mathbb{P}\left[\vec{U}_m \cap \mathcal{T}_j^{(N)} = \varnothing\right] \\
&= 1 - \mathbb{E}\left[\frac{(N - |\mathcal{T}_j^{(N)}|) \cdots (N+1-m-|\mathcal{T}_j^{(N)}|)}{N \cdots (N+1-m)}\right] \\
&= 1 - \frac{(N-m)!}{N!} \mathbb{E}\left[(N - |\mathcal{T}_j^{(N)}|) \cdots (N+1-m-|\mathcal{T}_j^{(N)}|)\right]
\end{aligned} \tag{7}$$

since the event $\vec{U}_m \cap \mathcal{T}_j^{(N)} = \varnothing$ requires that each of the uniformly chosen multicast member nodes U_i, for $i = 1, \ldots, m$, should not lie in $\mathcal{T}_j^{(N)}$. Therefore, the mean of the random variable $B_n(m)$ follows from (3) with $\mathbb{P}[A_j \leq n] = \frac{n}{j}$ as

$$\begin{aligned}
\mathbb{E}[B_n(m)] &= \sum_{j=n}^{N} \mathbb{E}\left[\mathbf{1}_{\{A_j \leq n\}}\right] \mathbb{P}[Z_j(m) > 0] = n \sum_{j=n}^{N} \frac{\mathbb{P}[Z_j(m) > 0]}{j} \\
&= n \sum_{j=n}^{N} \frac{1}{j} \left(1 - \frac{(N-m)!}{N!} \mathbb{E}\left[(N - |\mathcal{T}_j^{(N)}|) \cdots (N+1-m-|\mathcal{T}_j^{(N)}|)\right]\right).
\end{aligned} \tag{8}$$

In [1], we have shown[1] that, for $j > 0$,

$$\mathbb{P}\left[|\mathcal{T}_j^{(N)}| = n\right] = \frac{j(N-j)!(N-n)!}{N!(N-j-n+1)!}.$$

[1] The number of nodes N and node label j in [1] should be replaced here by $N+1$ and by $j+1$.

Since j is never equal to the root 0, it means that the largest subtree tree $|\mathcal{T}_j^{(N)}|$ is of size N, but never smaller than $|\mathcal{T}_j^{(N)}| = 1$ (namely the node j itself). The probability generating function $\varphi_{|\mathcal{T}_j^{(N)}|}(z)$ of $|\mathcal{T}_j^{(N)}|$ is

$$\varphi_{|\mathcal{T}_j^{(N)}|}(z) = \mathbb{E}\left[z^{|\mathcal{T}_j^{(N)}|}\right] = \sum_{n=1}^{N} \mathbb{P}\left[|\mathcal{T}_j^{(N)}| = n\right] z^n = \frac{j(N-j)!}{N!} \sum_{n=1}^{N} \frac{(N-n)!}{(N-j-n+1)!} z^n$$
$$= \frac{j(N-j)!}{N!} \sum_{k=0}^{N-1} \frac{k!}{(k-(j-1))!} z^{N-k}$$

Thus,

$$\mathbb{E}\left[(N-|\mathcal{T}_j^{(N)}|)\cdots(N+1-m-|\mathcal{T}_j^{(N)}|)\right] = \frac{d^m}{dz^m} E\left[z^{N-|\mathcal{T}_j^{(N)}|}\right]\Big|_{z=1} = \frac{d^m}{dz^m} E\left[z^N \varphi_{|\mathcal{T}_j^{(N)}|}\left(z^{-1}\right)\right]\Big|_{z=1}$$
$$= \frac{j(N-j)!}{N!} \sum_{k=0}^{N-1} \frac{k!}{(k-(j-1))!} \frac{d^m}{dz^m} E\left[z^k\right]\Big|_{z=1}$$
$$= \frac{j(N-j)!}{N!} \sum_{k=0}^{N-1} \frac{k!}{(k-(j-1))!} \frac{k!}{(k-m)!}$$

Substituted in (8) gives

$$\mathbb{E}\left[B_n(m)\right] = \sum_{j=n}^{N} \frac{n}{j} \left(1 - \frac{(N-m)!j(N-j)!}{(N!)^2} \sum_{k=0}^{N-1} \frac{(k!)^2}{(k-(j-1))!(k-m)!}\right)$$

and in (5)

$$\mathbb{E}[W_N(m)] = \sum_{n=1}^{N} \frac{1}{N+1-n} \sum_{j=n}^{N} \frac{1}{j} \left(1 - \frac{(N-m)!j(N-j)!}{(N!)^2} \sum_{k=0}^{N-1} \frac{(k!)^2}{(k-(j-1))!(k-m)!}\right)$$

The first sum is, after reversal of the summations and using the identity $\sum_{j=1}^{N} \frac{1}{j} \sum_{k=N+1-j}^{N} \frac{1}{k} = \sum_{k=1}^{N} \frac{1}{k^2}$ proved [3, Appendix], equal to

$$\sum_{n=1}^{N} \frac{1}{(N+1-n)} \sum_{j=n}^{N} \frac{1}{j} = \sum_{k=1}^{N} \frac{1}{k^2}$$

The second sum is

$$Y_m = \frac{(N-m)!}{(N!)^2} \sum_{n=1}^{N} \frac{1}{N+1-n} \sum_{j=n}^{N} (N-j)! \sum_{k=0}^{N-1} \frac{(k!)^2}{(k-(j-1))!(k-m)!}$$
$$= \frac{(N-m)!}{(N!)^2} \sum_{n=1}^{N} \frac{1}{N+1-n} \sum_{k=0}^{N-1} \frac{(k!)^2}{(k-m)!} \sum_{j=n}^{N} \frac{(N-j)!}{(k-(j-1))!}$$

Application of the identity

$$\sum_{j=n}^{m} \frac{(a-j)!}{(b-j)!} = \frac{1}{a+1-b} \left\{ \frac{(a-n+1)!}{(b-n)!} - \frac{(a-m)!}{(b-m-1)!} \right\} \tag{9}$$

gives

$$
\begin{aligned}
Y_m &= \frac{(N-m)!}{(N!)^2}\sum_{n=1}^{N}\sum_{k=0}^{N-1}\frac{(k!)^2\,(N-n)!(N-k-1)!}{(k-m)!(k+1-n)!(N-k)!} \\
&= \frac{(N-m)!}{(N!)^2}\sum_{k=0}^{N-1}\frac{(k!)^2}{(k-m)!(N-k)}\sum_{n=1}^{N}\frac{(N-n)!}{(k+1-n)!}
\end{aligned}
$$

Again, using (9) yields

$$
\begin{aligned}
Y_m &= \frac{(N-m)!}{N!}\sum_{k=0}^{N-1}\frac{k!}{(k-m)!(N-k)^2} \\
&= \frac{(N-m)!}{N!}\sum_{k=1}^{N}\frac{(N-k)!}{(N-m-k)!k^2} = \frac{1}{\binom{N}{m}}\sum_{k=1}^{N}\binom{N-k}{m}\frac{1}{k^2}
\end{aligned}
$$

Hence, we arrive at

$$
\mathbb{E}[W_N(m)] = \sum_{k=1}^{N}\frac{1}{k^2} - Y = \sum_{k=1}^{N}\frac{\binom{N}{m}-\binom{N-k}{m}}{\binom{N}{m}}\frac{1}{k^2} \tag{10}
$$

$$
= \sum_{j=1}^{m}\frac{1}{N+1-j}\sum_{k=j}^{N}\frac{1}{k} \tag{11}
$$

where the last formula (11) was found in [1]. Equality of both formulae is proved in Appendix A.

4 Acknowledgements

I would like to thank G. Hooghiemstra and R. van der Hofstad for their great interactions in our study of the stochastic shortest path problem.

References

[1] R. van der Hofstad, G. Hooghiemstra, and P. Van Mieghem. Size and weight of shortest path trees with exponential link weights. *Combinatorics, Probability and Computing*, vol. 15, pp. 903-926, 2006.

[2] P. Van Mieghem, G. Hooghiemstra, and R. van der Hofstad. On the Efficiency of Multicast. *IEEE/ACM Transactions on Networking*, Vol. 9, No. 6, pp. 719-732, 2001.

[3] R. van der Hofstad, G. Hooghiemstra, and P. Van Mieghem. The Weight of the Shortest Path Tree. *Random Structures and Algorithms*, Random Structures and Algorithms, Vol. 30, No. 3, May, pp. 359-379, 2007.

[4] P. Van Mieghem and S. Tang, "Weight of the Shortest Path to the First Encountered Peer in a Peer Group of Size m", accepted for Probability in the Engineering and Informational Sciences (PEIS), 2008.

[5] P. Van Mieghem. Performance Analysis of Communications Systems and Networks. Cambridge University Press, 2006.

A Proof that (10) equals (11)

Writing the difference in Y as $\Delta Y = Y_m - Y_{m-1} = -\Delta W$,

$$\Delta Y = \frac{1}{N!}\sum_{k=1}^{N}\left[\frac{(N-m)!}{(N-m-k)!} - \frac{(N-m+1)!}{(N-m+1-k)!}\right]\frac{(N-k)!}{k^2}$$

and with

$$B = \frac{(N-m)!}{(N-m-k)!} - \frac{(N-m+1)!}{(N-m+1-k)!} = \frac{-(N-m)!k}{(N-m+1-k)!}$$

we see that

$$\begin{aligned}\Delta Y &= \frac{-(N-m)!}{N!}\sum_{k=1}^{N}\frac{(N-k)!}{(N-m+1-k)!}\frac{1}{k}\\ &= \frac{-(N-m)!}{N!}\sum_{k=1}^{N-m+1}\frac{(N-k)!}{(N-m+1-k)!}\frac{1}{k}\end{aligned}$$

Using the identity

$$\sum_{j=n}^{b}\frac{1}{j}\frac{(a-j)!}{(b-j)!} = \frac{a!}{b!}\sum_{j=n}^{b}\frac{1}{j} - \frac{a!}{(b-n)!}\sum_{q=0}^{a-b-1}\frac{1}{a-b-q}\frac{(a-q-n)!}{(a-q)!}$$

which is valid for all integers a, b and n, gives

$$\sum_{j=1}^{N-m+1}\frac{1}{j}\frac{(N-j)!}{(N-m+1-j)!} = \frac{N!}{(N-m+1)!}\sum_{j=1}^{N-m+1}\frac{1}{j} - \frac{N!}{(N-m)!}\sum_{q=0}^{m-2}\frac{1}{m-1-q}\frac{1}{(N-q)}$$

Since

$$\begin{aligned}\sum_{q=0}^{m-2}\frac{1}{m-1-q}\frac{1}{(N-q)} &= \frac{1}{N-m+1}\sum_{q=0}^{m-2}\frac{1}{m-1-q} - \frac{1}{N-m+1}\sum_{q=0}^{m-2}\frac{1}{(N-q)}\\ &= \frac{1}{N-m+1}\sum_{j=1}^{m-1}\frac{1}{j} - \frac{1}{N-m+1}\sum_{j=N-m+2}^{N}\frac{1}{j}\end{aligned}$$

we have

$$\begin{aligned}\Delta Y &= -\frac{1}{(N-m+1)}\sum_{j=1}^{N-m+1}\frac{1}{j} + \frac{1}{N-m+1}\sum_{j=1}^{m-1}\frac{1}{j} - \frac{1}{N-m+1}\sum_{j=N-m+2}^{N}\frac{1}{j}\\ &= -\frac{1}{N-m+1}\sum_{j=m}^{N}\frac{1}{j}\end{aligned}$$

From the original expression (11), we immediately find that $\Delta W = \frac{1}{N+1-m}\sum_{k=m}^{N}\frac{1}{k}$ which proves equality in the differences since $\Delta Y = -\Delta W$. Equality of (11) and (10) then follows since for $m = N$, both expressions are equal. $\square$