

On Tries, Contention Trees and Their Analysis

Stephan Wagner

Department of Mathematical Sciences, Stellenbosch University, Private Bag X1, Matieland 7602,
South Africa
swagner@sun.ac.za

Received May 15, 2007

AMS Subject Classification: 68W40; 94A40, 68P05

Abstract. In this paper a survey on tries, a contention resolution algorithm, their similarities, dissimilarities, and their mathematical treatment, will be given. It has already been mentioned in some papers that tries and contention trees follow one common stochastic model, but still they are frequently treated as separate objects in the literature. Hence the aim of the current work is to contribute to the unification of the various results in that area and to exhibit the employed methods, which involve, among others, analytic poissonization/depoissonization and the Mellin transform. For the sake of the example, a new parameter in contention trees, the number of *terminal frames*, will be studied.

Keywords: tries, contention trees, poissonization, Mellin transform

1. Introduction

1.1. Tries

We will treat two different kinds of combinatorial objects (which turn out to be almost isomorphic) in this paper, the first of which are *tries*. Tries are a very popular data structure for words over a finite alphabet, and they have also been treated quite extensively from a mathematical point of view. Let us start with a brief description.

Given a set of n strings over a finite alphabet with m letters (the most typical case being $m = 2$), a trie is built up in the following way: A string is stored in an external node, and a path from the root to an external node represents the shortest prefix of the string that is not a prefix of any other string. Each outgoing edge at an internal node represents one letter of the alphabet, so a trie is a special m -ary tree. Figure 1 shows a very simple example of a binary trie for the strings 000101, 110110, 001010, and 101011. An important special case is given by the *suffix trie*, which is built up from the suffixes over a single string.

Tries were first proposed by de la Briandais [3] in the context of information processing; the name was suggested by Fredkin [12], being part of the word *retrieval*. Tries are used in various applications in computer science due to their simplicity and

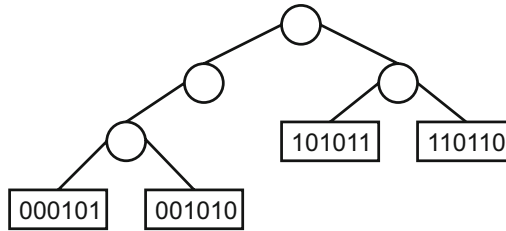


Figure 1: A binary trie.

efficiency. Examples include searching and sorting, fast retrieval (dynamic hashing, see [4, 21]), the IP-addresses lookup problem [30], polynomial factorization [22], data compression [29], and many others. The trie structure is also related to coin-flipping processes, e.g., the leader election algorithm discussed in [6, 28] (so-called “incomplete tries” — only one branch is pursued).

1.2. Contention Trees

The fundamental problem of multi-user communication is to serve many different senders, given only a single communication channel. For this purpose, the classical solution of time-division multiplexing becomes rather inefficient when there is a large number of senders and each of them is inactive most of the time. A far more efficient and elegant solution to this problem was given by the ALOHA protocol developed in the early 1970’s (cf. Abramson [33]), which provided immediate random access to the channel. If two transmitters are active at the same time, a *collision* occurs, and both transmitters try again later after some random time period.

The main disadvantage of the ALOHA protocol is its rather poor performance if the channel occupancy is above a certain level — in fact, the protocol is unstable: If the senders submit their messages according to a Poisson process, then the backlog tends to ∞ with probability 1. One strategy to deal with this problem is the contention tree algorithm due to Capetanakis [2], Tsybakov and Mikhailov [33]. Suppose that a large number n of submitters contending for channel access is given; each of them receives ternary feedback on the outcome during a particular contention slot: Either zero (empty slot), one (successful transmission), or more than one (collision) transmitters have been broadcasting during this slot.

For some given nodal degree m , we let m consecutive slots be grouped into a contention frame. During the first contention frame, each transmitter picks a slot at random — usually, the probabilities are equal for all slots, but they may also be given by values p_1, \dots, p_m . If a collision occurs at some slot, a new frame is opened for the transmitters who made use of that slot. This procedure is repeated recursively (the algorithm is applied to each of the new frames, one after the other), inducing a tree structure as shown in Figure 2 (with 13 contenders and $m = 3$).

There are several variants of the contention tree algorithm, such as the contention stack algorithm due to Tsybakov and Vvedenskaya [34] and the Gallager-Tsybakov-Mikhailov algorithm, where the partition of the contenders is not based on a random

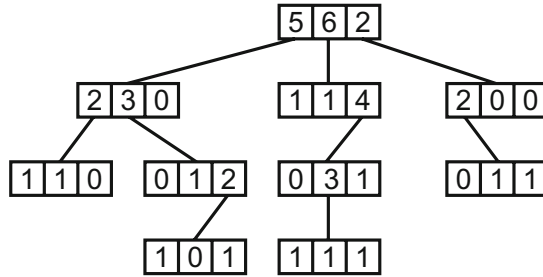


Figure 2: A contention tree.

process, but on the time at which a user became active [13].

In many papers, the contention tree algorithm is mentioned to be an application of the trie structure [15, 16, 25] (which is plain to see and will be exhibited in the following section); however, the algorithm is still frequently treated separately, without making use of the existing results for tries. One major aim of this paper is to enlarge the general awareness of the fact that tries and contention trees are two manifestations of one common underlying model.

2. Common Parameters, Similarities, and Dissimilarities

Now, we want to exhibit the fact that the contention tree algorithm and tries follow one common stochastic model. Indeed, we can identify each contender in our contention resolution algorithm with a word: Simply take the sequence of slots that is chosen by the contender (which is of course a word over an alphabet of m letters). Then, the position in the contention tree where the contender is finally able to successfully submit their data is exactly the position in the trie where the associated word is stored. In Figure 3, an example of a ternary contention tree and its associated trie is shown. Note that a trie is an m -ary tree whose leaves are regarded as *external nodes*, where the data are stored, whereas a contention tree only consists of internal nodes (it would be possible, though, to associate an additional external node to each contender). This fact can result in some minor differences for parameters in the two models.

In view of the described correspondence, a common stochastic model can be applied to tries and contention trees: We assume that a contender picks the i -th slot with a probability p_i , or equivalently that the i -th letter of our alphabet appears with probability p_i ($\sum_{i=1}^m p_i = 1$). This is known as the *Bernoulli model*. As already mentioned, it is usually reasonable to assume that $p_1 = p_2 = \dots = p_m = \frac{1}{m}$ (the so-called *symmetric case*) in the contention tree algorithm, since the purpose of the algorithm is to split the group of contenders as effectively as possible. However, there are slightly modified algorithms which are more efficient in the asymmetric case, and it is also quite plausible that the frequencies of the letters of our alphabet are all different. Yet the symmetric case is generally easier to analyze (see, for instance, [20], where the specific difficulties of the asymmetric case are explained), and there are also parameters where

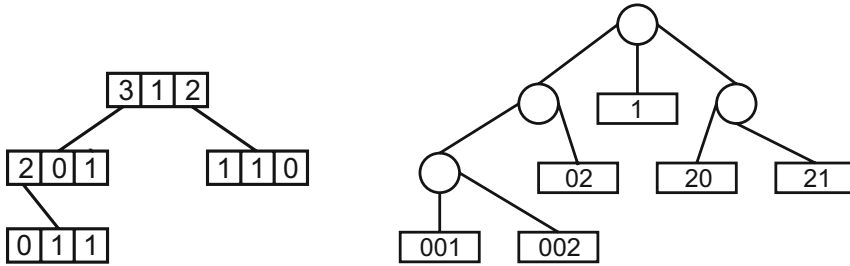


Figure 3: A contention tree and its associated trie.

the behaviors differ in the symmetric and the asymmetric cases.

The case $m = 2$ is certainly the most interesting and important one, especially in the case of tries in view of their applications in computer science. Hence, many papers deal exclusively with this special case, but there are also generalizations to the general m -ary case (see, [32], for instance).

There is a great variety of parameters that are of interest for tries and/or contention trees. Among them, the most important parameter for the performance of contention trees is probably the number of contention frames (equivalently, the size of a trie), whose average behavior has been studied in several papers. The following formula was given by Janssen and de Jong [18]:

Theorem 2.1. (Janssen and de Jong [18]) *If L_n denotes the average number of frames for a given number n of transmitters, the asymptotic behavior of L_n for $n \rightarrow \infty$ is given by*

$$\frac{L_n}{n} = \frac{1}{\log m} + \Phi_m(\log_m n) + O(1), \tag{2.1}$$

where Φ_m is a periodic function (of very small amplitude).

Their proof used an explicit sum formula for L_n :

$$L_n = \sum_{d=0}^{\infty} m^d \left(1 - \left(1 - m^{-d} \right)^n \right) - n \left(1 - m^{-d} \right)^{n-1},$$

whose asymptotics can be determined by means of the Poisson sum formula (in Section 3, we will perform a similar analysis, but employ the Mellin transform for this purpose). However, the asymptotic behavior shown in formula (2.1) had already been given before by Flajolet and Jacquet [7] (in the case $m = 2$), and they attributed the analysis to Knuth [21].

Another common approach for problems of this kind is analytic poissonization and de-poissonization: One considers the Poisson transform $L(x) = e^{-x} \sum_{n=0}^{\infty} \frac{L_n}{n!} x^n$, which satisfies a nice functional equation:

$$L(mx) = mL(x) + 1 - (1 + mx)e^{-mx}.$$

It will be shown in Section 3 how equations of this kind can be derived. Now it follows from general theorems (see the excellent survey article of Jacquet and Szpankowski

[17]) that $L_n \sim L(n)$, which allows one to determine the asymptotic behavior. Indeed, in a recent paper, Györfi and Györfi [14] investigated the asymptotic difference between L_n and its Poisson transform $L(n)$ in detail to obtain more precise information on the behavior of L_n depending on n .

However, even more is known — Jacquet and Régnier [16] have already shown in 1988 (in the case of symmetric or unsymmetric binary tries) that the distribution of the size of a trie tends to a normal distribution, and that the moments converge to the corresponding moments for the normal distribution:

Theorem 2.2. (Jacquet and Régnier [16]) *If S_n is the (random) size of a trie built from n strings, then the distribution of the normalized random variable tends to a normal distribution:*

$$\frac{S_n - \mathbb{E}(S_n)}{\sqrt{\mathbb{V}(S_n)}} \xrightarrow{d} N(0, 1).$$

The moments also converge, and the convergence rate is $O(n^{\varepsilon-1/2})$ for arbitrary $\varepsilon > 0$.

Another parameter of interest is the average number of levels (attempts) a random contender requires for successful transmission — this parameter is also known as the *depth* (distance from the root to a randomly selected external node, see [15] for instance) and is related to the *external path length* (the sum of all distances from the external nodes to the root). In the case of tries, this gives the average time for searching a word in the trie. However, it is not true that the average number of levels a contender needs is also a measure for the average time that a contender has to wait until the message can be sent: Note that all frames to the left of a certain frame in a contention tree are executed before it. Hence, in order to determine the average waiting time, one needs another parameter that has no immediate interpretation for tries. It seems that, unlike the well-studied depth, the average waiting time of a contender has not been investigated as extensively yet; its analysis should follow the same lines though. For the average number of attempts that a contender requires, Janssen and de Jong provide the formula

$$d_n = \log_m n + \frac{1}{2} + \frac{\gamma}{\log m} + \Psi_m(\log_m n) + O(n^{-1}),$$

where Ψ_m is a periodic function and γ denotes the Euler-Mascheroni constant. Again, limiting distributions (for the depth, but apparently not for the average waiting time, see above) have been studied in the binary case, see [15, 27].

A very similar parameter is the number of tree levels required to complete the whole algorithm (i.e., the largest number of levels any contender needs). Again, this should not be confused with the time the algorithm takes. In the language of trees (and, in particular, tries), this is known as the *height* (longest path from the root to a leaf).

These are just three examples of important parameters in the analysis of tries and contention trees. As Hwang, Nicodème, Park and Szpankowski [25, 26] pointed out, these and other important parameters can be expressed in terms of the *profile* of tries. The internal/external profile of the tree is defined as the total number of internal/external nodes at a given level, i.e., the random variables $I_{n,k}$ and $B_{n,k}$ are given by the number of internal/external nodes at level k in a random trie (where the root is at level 0). Important parameters that can be expressed in terms of the profile include, apart from the aforementioned,

- (internal) path length, i.e., the sum of the distances between the internal nodes and the root, which is given by $\sum_j jI_{n,j}$,
- shortest path from the root to an external node: $\min \{j: B_{n,j} > 0\}$, and
- fill-up level (largest full level): $\max \{j: I_{n,j} = m^j\}$.

Hence it is not exaggerated to state that the profile provides an almost complete characterization of the shape of a trie. The main result for the profile, which follows from a quite intricate analysis, can be roughly stated as follows: The distributions of the internal and external profile converge (in distribution) to a Gaussian distribution, provided that the variance tends to infinity (which is the case if $\lim_{n \rightarrow \infty} \frac{k}{\log n}$ lies within a certain range); otherwise, it tends to a Poisson distribution or to 0 (the latter being the case for very large or small k ; the former case occurs if the variance is of order $\Theta(1)$, i.e., bounded above and below by constants). For details on the distributions and on the asymptotics of the mean and variance, see the papers [25] and [26].

3. Analysis of Terminal Frames

In order to briefly demonstrate some of the techniques that are applied in the analysis of parameters of contention trees and tries, we investigate the *number of terminal frames* in a contention tree; this gives us information about the proportion of those frames that are merely used for splitting purposes compared to those in which the transmission actually occurs. There are two plausible ways to define “terminal”: We call a contention frame

- terminal of the first type if all contenders of the frame can transmit successfully; equivalently, no splitting occurs at such a frame any more,
- terminal of the second type if at least one successful transmission is performed at this frame.

In terms of tries, these are the number of internal nodes whose children are all external nodes and the number of internal nodes with at least one external node among the children respectively. We are going to provide explicit formulas for the average number of terminal frames of both types as well as asymptotic expressions — not surprisingly, oscillation phenomena (as for the size L_n of contention trees, for instance) can be observed.

3.1. Explicit Formulas

Proposition 3.1. *The average number of terminal frames of the first type, given the number n of contenders, is*

$$t_n = \sum_{d=0}^{\infty} m^d \sum_{k=2}^n \binom{n}{k} m^{-dk} (1 - m^{-d})^{n-k} (m)_k m^{-k}.$$

The average number of terminal frames of the second type, given the number n of contenders, is

$$T_n = n - \sum_{d=0}^{\infty} m^d \sum_{l=2}^m (-1)^l (m)_l \binom{n}{l} m^{-(d+1)l} (1 - lm^{-(d+1)})^{n-l}.$$

Proof. We apply the probabilistic model that is also used by Janssen and de Jong [18]: We think of the contention tree as an infinite, complete m -ary tree. At the first level, each contender chooses one of the m slots at random and proceeds to the corresponding node of the next level; this process is repeated for each subsequent level. Therefore, the contenders are randomly distributed among the m^d nodes of the d -th level. The probability that k of n contenders are assigned to some level- d node is therefore given by

$$\binom{n}{k} m^{-dk} (1 - m^{-d})^{n-k}.$$

The first time contenders pick a slot which they don't have to share is the time when their transmission is successfully completed.

Now consider t_n first. Obviously, the probability that a contention frame with k contenders is a terminal frame of the first type is given by

$$(m)_k m^{-k},$$

where $(m)_k = m(m-1) \cdots (m-k+1)$ denotes the falling factorial, since there are $(m)_k$ among the m^k possibilities for which all contenders occupy their own slots. Thus, we have a formula for the probability that a frame at level d is a terminal frame of the first type:

$$\sum_{k=2}^n \binom{n}{k} m^{-dk} (1 - m^{-d})^{n-k} (m)_k m^{-k}.$$

Note that we have to sum over $k \geq 2$ since a contention frame is only generated when there are at least two colliding contenders. Summing over all frames now yields the formula for t_n . In a similar manner, we note that the probability that at least l ($l \geq 2$) contenders have successful transmission at a certain level- d frame is given by

$$\binom{n}{l} (m)_l m^{-(d+1)l} (1 - l m^{-(d+1)})^{n-l}.$$

By the inclusion-exclusion principle, we obtain the formula for T_n . ■

The presented approach works well if one is only interested in expected values. However, if higher moments or limit distributions are requested, it is much more advantageous to use a recursive approach and generating functions instead, as exhibited in the following section and parts of Section 3.3.

3.2. Functional Equations and the Special Case $m = 2$

Each contender transmits successfully at some terminal frame of the second type. If this contender is not the only one who transmits at this frame, then it is also a terminal frame of the first type if $m = 2$. In this case there are two slots of the frame occupied by single contenders. This shows that the total number of contenders must equal the sum of the total number of terminal frames of the first kind and the total number of terminal frames of the second kind, i.e.,

$$t_n + T_n = n,$$

if $m = 2$. This is not the only interesting feature of the case $m = 2$. We are going to show that the exponential generating functions of t_n and T_n satisfy specifically nice functional equations. First, we note that the formulas of Proposition 3.1 reduce to

$$t_n = \sum_{d=0}^{\infty} 2^{-d-1} \binom{n}{2} (1 - 2^{-d})^{n-2}$$

and $T_n = n - t_n$. Now, we define $t_0 = t_1 = T_0 = 0$ and $T_1 = 1$ and let $t(x) = e^{-x} \sum_{n=0}^{\infty} \frac{t_n}{n!} x^n$ and $T(x) = e^{-x} \sum_{n=0}^{\infty} \frac{T_n}{n!} x^n$ be the Poisson transforms of t_n and T_n respectively. From the identity $t_n + T_n = n$, it is clear that $t(x) + T(x) = x$. Now, we insert the formula for t_n to obtain

$$\begin{aligned} t(x) &= e^{-x} \sum_{n=2}^{\infty} \sum_{d=0}^{\infty} 2^{-d-1} \binom{n}{2} (1 - 2^{-d})^{n-2} \frac{x^n}{n!} \\ &= x^2 e^{-x} \sum_{n=2}^{\infty} \sum_{d=0}^{\infty} 2^{-d-2} \frac{((1 - 2^{-d})x)^{n-2}}{(n-2)!} \\ &= \frac{x^2 e^{-x}}{4} \sum_{d=0}^{\infty} 2^{-d} \exp\left(\left(1 - 2^{-d}\right)x\right) \\ &= \frac{x^2}{4} \sum_{d=0}^{\infty} 2^{-d} \exp\left(-2^{-d}x\right). \end{aligned}$$

Now, replacing x by $2x$ yields

$$\begin{aligned} t(2x) &= x^2 \sum_{d=0}^{\infty} 2^{-d} \exp\left(-2^{-d+1}x\right) \\ &= x^2 \sum_{d=-1}^{\infty} 2^{-d-1} \exp\left(-2^{-d}x\right) \\ &= x^2 e^{-2x} + \frac{x^2}{2} \sum_{d=0}^{\infty} 2^{-d} \exp\left(-2^{-d}x\right) \\ &= x^2 e^{-2x} + 2t(x). \end{aligned}$$

As a corollary, we have $T(2x) = -x^2 e^{-2x} + 2T(x)$. These functional equations reflect the recursive structure of contention trees. A simple combinatorial argument shows that we have

$$t_n = 2^{-n} \sum_{k=0}^n \binom{n}{k} (t_k + t_{n-k}) + \frac{\delta_{n2}}{2},$$

from which the functional equation follows easily as well. Indeed, the probability that the n contenders split into one group of k contenders and another group of $n - k$ contenders in the first step is exactly $\binom{n}{k} 2^{-n}$; afterwards, we have an average of $t_k + t_{n-k}$ terminal frames of the first type, except for the case $n = 2$, where we have to add 1

with probability $\frac{1}{2}$. Similar arguments show that in the general case, there is always a recursion of the form

$$t_n = m^{-n} \sum_{k_1+k_2+\dots+k_m=n} \binom{n}{k_1, k_2, \dots, k_m} (t_{k_1} + t_{k_2} + \dots + t_{k_m}) + r_n$$

and thus an analytic function $f_m(x)$ such that

$$t(mx) = mt(x) + f_m(x).$$

This kind of linear additive functional equation is typical for the probabilistic models of tries, contention trees, and many similar structures, and it has been studied extensively (see for example, [5, 31]). The order of the steps can be reversed — frequently, explicit formulas are deduced from functional equations by iteration, e.g.,

$$t(x) = \frac{x^2 e^{-x}}{4} + 2t\left(\frac{x}{2}\right) = \frac{x^2 e^{-x}}{4} + \frac{x^2 e^{-x/2}}{8} + 4t\left(\frac{x}{4}\right) + \dots$$

The behavior of the sum $\sum_{d=0}^{\infty} 2^{-d} \exp(-2^{-d}x)$ that appears in the explicit formula for $t(x)$ can be determined by means of the Mellin transform (see the papers of Flajolet et al. [8–11] for various applications): The Mellin transform is always particularly useful when one is dealing with sums of the type $g(x) = \sum_{n \geq 1} \lambda_n f(\mu_n x)$, since the Mellin transforms $f^*(s) = \int_0^{\infty} x^{s-1} f(x) dx$ and $g^*(s)$ are connected via

$$g^*(s) = \sum_{n \geq 1} \lambda_n \mu_n^{-s} f^*(s).$$

In our case, the Mellin transform of e^{-x} is given by $\Gamma(s)$, and thus the transform of the sum we are interested in is

$$\sum_{d=0}^{\infty} 2^{-d} 2^{ds} \Gamma(s) = \frac{\Gamma(s)}{1 - 2^{s-1}}.$$

Using the Mellin inversion formula, we have

$$\sum_{d=0}^{\infty} 2^{-d} \exp(-2^{-d}x) = \frac{1}{2\pi i} \int_{1/2-i\infty}^{1/2+i\infty} \frac{\Gamma(s)}{1 - 2^{s-1}} x^{-s} ds.$$

Shifting the path of integration to the right and collecting residues at the points $1 + \frac{2\pi ia}{\log 2}$ ($a \in \mathbb{Z}$), we obtain

$$\sum_{d=0}^{\infty} 2^{-d} \exp(-2^{-d}x) = \frac{1}{x \log 2} \sum_{a=-\infty}^{\infty} \exp(-2\pi ia \log_2 x) \Gamma\left(1 + \frac{2\pi ia}{\log 2}\right) + O(x^{-2}).$$

Finally, from general depoissonization theorems [17], we find that

$$t_n = t(n) + o(n) = \frac{n}{4 \log 2} \sum_{a=-\infty}^{\infty} \exp(-2\pi ia \log_2 n) \Gamma\left(1 + \frac{2\pi ia}{\log 2}\right) + o(n).$$

The function that is represented by the Fourier series

$$F(x) = \sum_{a=-\infty}^{\infty} \Gamma\left(1 + \frac{2\pi ia}{\log 2}\right) \exp(-2\pi i ax)$$

is periodic, and the non-constant terms yield a function of very small amplitude — compare the formula of Janssen and de Jong given in Theorem 2.1.

3.3. Asymptotic Evaluation

In a similar way, the Mellin transform can be used to determine the asymptotic behavior of the quantities t_n and T_n from the explicit formulas: For this purpose, we study the infinite sum

$$\sum_{d=0}^{\infty} m^{-(k-1)d} \left(1 - \alpha m^{-d}\right)^{n-k},$$

where $0 < \alpha \leq 1$ and k, m are fixed. We are going to show that it behaves, apart from an oscillatory term, like $\frac{\Gamma(k-1)}{\log m} (\alpha(n-k))^{-(k-1)}$. For this purpose, we replace $1 - \alpha m^{-d}$ by $\exp(-\alpha m^{-d})$ and estimate the error. Fix some positive $\varepsilon < \frac{1}{k+1}$. Then, in the case that $d \leq (1 - \varepsilon) \log_m(n)$, we have $m^{-d} \geq n^{\varepsilon-1}$ and thus

$$\left(1 - \alpha m^{-d}\right)^{n-k} \leq \exp\left(-\alpha m^{-d}(n-k)\right) \leq \exp\left(-\alpha n^{\varepsilon-1}(n-k)\right),$$

which shows that the contribution of the terms with $d \leq (1 - \varepsilon) \log_m(n)$ decreases with $\exp(-n^\varepsilon)$. On the other hand, we make use of the inequality $\exp(-x) - (1-x) \leq \frac{x^2}{2}$ for positive x to show that

$$\begin{aligned} & \sum_{d > (1-\varepsilon) \log_m(n)} m^{-(k-1)d} \left(\exp\left(-\alpha(n-k)m^{-d}\right) - \left(1 - \alpha m^{-d}\right)^{n-k} \right) \\ & \leq \sum_{d > (1-\varepsilon) \log_m(n)} m^{-(k-1)d} (n-k) \left(\exp\left(-\alpha m^{-d}\right) - \left(1 - \alpha m^{-d}\right) \right) \\ & \leq \sum_{d > (1-\varepsilon) \log_m(n)} m^{-(k-1)d} (n-k) \frac{\alpha^2 m^{-2d}}{2} \\ & \ll nm^{-(k+1)(1-\varepsilon) \log_m(n)} \\ & = n^{-k+\varepsilon(k+1)}. \end{aligned}$$

Now we consider the behavior of

$$\sum_{d=0}^{\infty} m^{-(k-1)d} \exp\left(-\alpha m^{-d}(n-k)\right).$$

Using the Mellin transform again (alternatively, the Poisson summation formula could also be applied), we obtain

$$\begin{aligned} & \sum_{d=0}^{\infty} m^{-(k-1)d} \exp\left(-\alpha m^{-d}(n-k)\right) \\ &= \frac{1}{\log m} (\alpha(n-k))^{-(k-1)} \sum_{a=-\infty}^{\infty} \exp(-2\pi i a \log_m(\alpha(n-k))) \Gamma\left(k-1 + \frac{2\pi i a}{\log m}\right) \\ & \quad + O\left(n^{-k}\right). \end{aligned}$$

Altogether, we obtain the following asymptotic formulas for t_n and T_n :

Proposition 3.2. *The average number of terminal frames of the first type is given by*

$$t_n = \frac{n}{\log m} \sum_{k=2}^m \binom{m}{k} m^{-k} \sum_{a=-\infty}^{\infty} \exp(-2\pi i a \log_m(n-k)) \Gamma\left(k-1 + \frac{2\pi i a}{\log m}\right) + o(n).$$

Note that only the summands up to $k = m$ are necessary, since $\binom{m}{k} = 0$ for $k > m$. Analogously, we have

Proposition 3.3. *The average number of terminal frames of the second type is given by*

$$\begin{aligned} T_n = n \left(1 - \frac{1}{m \log m} \sum_{l=2}^m \frac{(-1)^l}{l^{l-1}} \binom{m}{l} \right. \\ \left. \sum_{a=-\infty}^{\infty} \exp(-2\pi i a \log_m(l(n-l))) \Gamma\left(l-1 + \frac{2\pi i a}{\log m}\right) \right) + o(n). \end{aligned}$$

Note again that the coefficients of n are not constant, but periodic functions in $\log_m(n)$.

As mentioned in Section 2, it is also possible to obtain limiting distributions for parameters like the size (number of frames); an example of such a result is stated in Theorem 2.2. In order to achieve this, one has to consider bivariate generating functions — in the case of terminal frames, let $t_{n,k}$ be the probability that a contention tree with n contenders has exactly k terminal frames of the first type and $T_{n,k}$ the probability that a contention tree with n contenders has exactly k terminal frames of the second type. In order to obtain a simpler functional equation, we set $t_{1,0} = T_{1,0} = 1$ and $t_{1,1} = T_{1,1} = 0$. Then, the relevant generating functions

$$s(x, z) := e^{-x} \sum_{n,k} t_{n,k} \frac{x^n}{n!} z^k$$

and

$$S(x, z) := e^{-x} \sum_{n,k} T_{n,k} \frac{x^n}{n!} z^k$$

satisfy, in view of the recursive structure,

$$s(x, z) = s\left(\frac{x}{m}, z\right)^m + (z - 1)e^{-x} \left(\left(\frac{x}{m} + 1\right)^m - 1 - x\right)$$

and

$$S(x, z) = zS\left(\frac{x}{m}, z\right)^m - (z - 1) \left(xe^{-x} + \left(S\left(\frac{x}{m}, z\right) - \frac{x}{m}e^{-x/m}\right)^m\right).$$

However, to keep the paper short, we won't perform any analysis with these bivariate functions.

Finally, we consider the coefficients of the non-oscillating parts in the asymptotic formulas for t_n and T_n , namely,

$$a(m) = \frac{1}{\log m} \sum_{k=2}^m \binom{m}{k} m^{-k} (k-2)!$$

and

$$b(m) = 1 - \frac{1}{m \log m} \sum_{l=2}^m \frac{(-1)^l}{l^{l-1}} \binom{m}{l} (l-2)!$$

and study their behavior when m is large. The first sum is easily estimated by elementary means: We divide the range of summation into the three parts $[2, m^{1/3}]$, $[m^{1/3}, m^{1/2+\epsilon}]$, and $[m^{1/2+\epsilon}, m]$. It is not difficult to see that the third part is smaller than any power of m . For the other parts, we apply Stirling's formula:

$$\begin{aligned} \binom{m}{k} m^{-k} (k-2)! &= \frac{1}{k(k-1)} \left(1 + \frac{k}{2m} + \frac{3k^2}{8m^2} + \dots\right) \\ &\quad \exp\left(-\frac{k^2}{2m} - \frac{k^3}{6m^2} - \frac{k^4}{12m^3} + \dots\right). \end{aligned}$$

For $k \leq m^{1/3}$, we may expand the exponential in another series and then calculate the sums; for $k \geq m^{1/3}$, we approximate the sum by an integral. Elementary calculations show that

$$a(m) = \frac{1}{\log m} \sum_{k=2}^m \binom{m}{k} m^{-k} (k-2)! = \frac{1}{\log m} \left(1 - \sqrt{\frac{\pi}{2m}} + \dots\right).$$

The second sum is alternating, so its asymptotic evaluation is somewhat trickier. We use the technique of "Rice's integrals" (cf. Flajolet and Sedgewick [11]) to write the sum as a complex integral:

$$\begin{aligned} \sum_{l=2}^m \frac{(-1)^l}{l^{l-1}} \binom{m}{l} (l-2)! &= \frac{(-1)^{m-1}}{2\pi i} \int_{3/2-i\infty}^{3/2+i\infty} \frac{m! \Gamma(s-1) s^{-s+1}}{s(s-1) \dots (s-m)} ds \\ &= \frac{(-1)^{m-1}}{2\pi i} \int_{3/2-i\infty}^{3/2+i\infty} \frac{m! \Gamma(s-m)}{s^s (s-1)} ds. \end{aligned}$$

Shifting the path of integration to the line $\text{Re } s = \frac{1}{2}$ and collecting the residue at 1 yields

$$\sum_{l=2}^m \frac{(-1)^l}{l^{l-1}} \binom{m}{l} (l-2)! = m(\log m - 1) - \frac{1}{2} + O\left(\frac{1}{m}\right) + \frac{(-1)^{m-1}}{2\pi i} \int_{1/2-i\infty}^{1/2+i\infty} \frac{m!\Gamma(s-m)}{s^s(s-1)} ds.$$

Now, we have to deform the contour to a curve surrounding 0 at a radius of $\frac{1}{\log m}$ (cf. [11]). Then the contribution of the logarithmic singularity around 0, whose expansion is given by

$$(-1)^{m-1} \frac{m!\Gamma(s-m)}{s^s(s-1)} = \frac{1}{s} - \log s + O(1),$$

is seen to be $1 + O\left(\frac{\log \log m}{\log m}\right)$; therefore, the second sum is

$$\sum_{l=2}^m \frac{(-1)^l}{l^{l-1}} \binom{m}{l} (l-2)! = m(\log m - 1) + \frac{1}{2} + \dots,$$

yielding

$$b(m) = \frac{1}{\log m} - \frac{1}{2m \log m} + \dots$$

So we see that in both cases the main term tends to $\frac{n}{\log m}$ for large m — which is the main asymptotic term for the average total number of frames; some specific values are given in the following table:

m	$a(m)$	$b(m)$	$\frac{1}{\log m}$
2	0.360674	0.639326	1.442695
3	0.337126	0.578593	0.910239
4	0.321225	0.533504	0.721348
5	0.309375	0.498516	0.621335
10	0.275333	0.398501	0.434294
20	0.245456	0.321580	0.333808
50	0.211946	0.252169	0.255622
100	0.190634	0.215736	0.217147

Acknowledgments. The author thanks Michael Drmota and Alois Panholzer for their valuable suggestions.

References

1. N. Abramson, Development of the alohanet, *IEEE Trans. Inform. Theory* **31** (2) (1985) 119–123.
2. J.I. Capetanakis, Tree algorithms for packet broadcast channels, *IEEE Trans. Inform. Theory* **25** (5) (1979) 505–515.

3. R. da la Briandais, File searching using variable length keys, In: Proceedings of the Western Joint Computer Conference, Vol. 15, New York, (1959) pp. 295–298.
4. R. Fagin, J. Nievergelt, N. Pippenger, and H. Raymond Strong, Extendible hashing — a fast access method for dynamic files, *ACM Trans. Database Syst.* **4** (3) (1979) 315–344.
5. G. Fayolle, P. Flajolet, and M. Hofri, On a functional equation arising in the analysis of a protocol for a multi-access broadcast channel, *Adv. in Appl. Probab.* **18** (2) (1986) 441–472.
6. J.A. Fill, H.M. Mahmoud, and W. Szpankowski, On the distribution for the duration of a randomized leader election algorithm, *Ann. Appl. Probab.* **6** (4) (1996) 1260–1283.
7. P. Flajolet and P. Jacquet, Analytic models for tree communication protocols, In: *Flow Control of Congested Networks*, Vol. 38, A.R. Odoni, L. Bianco, and G. Szegö, Eds., Springer-Verlag, (1987) pp. 223–234.
8. P. Flajolet and M. Golin, Mellin transforms and asymptotics: the mergesort recurrence, *Acta Inform.* **31** (7) (1994) 673–696.
9. P. Flajolet, X. Gourdon, and P. Dumas, Mellin transforms and asymptotics: harmonic sums, *Theoret. Comput. Sci.* **144** (1-2) (1995) 3–58.
10. P. Flajolet, P. Grabner, P. Kirschenhofer, H. Prodinger, and R.F. Tichy, Mellin transforms and asymptotics: digital sums, *Theoret. Comput. Sci.* **123** (2) (1994) 291–314.
11. P. Flajolet and R. Sedgewick, Mellin transforms and asymptotics: finite differences and Rice’s integrals, *Theoret. Comput. Sci.* **144** (1-2) (1995) 101–124.
12. E. Fredkin, Trie memory, *Comm. ACM* **3** (9) (1960) 490–499.
13. R.G. Gallagher, Conflict resolution in random access broadcast networks, In: *Proceedings of the AFOSR Workshop in Communication Theory and Applications*, Provincetown, MA, (1978) pp. 74–76.
14. L. Györfi and S. Györi, Analysis of tree algorithm for collision resolution, In: *2005 International Conference on Analysis of Algorithms*, Barcelona, (2005) pp. 357–364.
15. P. Jacquet and M. Régnier, Trie partitioning process: limiting distributions, In: *Lecture Notes in Comput. Sci.* Vol. 214, Springer, Berlin, (1986) pp. 196–211.
16. P. Jacquet and M. Régnier, Normal limiting distribution of the size of tries, In: *Performance’87*, North-Holland, Amsterdam, (1988) pp. 209–223.
17. P. Jacquet and W. Szpankowski, Analytical de-Poissonization and its applications, *Theoret. Comput. Sci.* **201** (1-2) (1998) 1–62.
18. A.J.E.M. Janssen and M.J.M. de Jong, Analysis of contention tree-algorithms, *IEEE Trans. Inform. Theory* **46** (6) (2000) 2163–2172.
19. M.A. Kaplan and E. Gulko, Analytic properties of multiple-access trees, *IEEE Trans. Inform. Theory* **31** (2) (1985) 255–263.
20. P. Kirschenhofer, H. Prodinger, and W. Szpankowski, On the variance of the external path length in a symmetric digital trie, *Discrete Appl. Math.* **25** (1-2) (1989) 129–143.
21. D.E. Knuth, *The Art of Computer Programming*, Vol. 3, Addison-Wesley Publishing Co., Reading, Mass., 1973.
22. D. Lazard, On polynomial factorization, In: *Proceedings of the European Computer Algebra Conference on Computer Algebra*, Lecture Notes in Comput. Sci., Vol. 44, Springer-Verlag, London, (1982) pp. 126–134.
23. J.L. Massey, Collision resolution algorithms and random access algorithms, In: *Multi-User Communication Systems*, Springer Verlag, New York, (1981) pp. 73–137.
24. P. Mathys and P. Flajolet, Q -ary collision resolution algorithms in random-access systems with free or blocked channel access, *IEEE Trans. Inform. Theory* **31** (2) (1985) 217–243.
25. G. Park, H.-K. Hwang, P. Nicodème, and W. Szpankowski, Profiles of tries, *SIAM J. Comput.*, to appear.
26. G. Park and W. Szpankowski, Towards a complete characterization of tries, In: *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, ACM, New York, (2005) pp. 33–42.

27. B. Pittel, Paths in a random digital tree: limiting distributions, *Adv. Appl. Probab.* **18** (1) (1986) 139–155.
28. H. Prodinger, How to select a loser, *Discrete Math.* **120** (1-3) (1993) 149–159.
29. R. Sedgewick, *Algorithms*, Addison-Wesley Publishing Co., Advanced Book Program, Reading, MA, 1983.
30. V. Srinivasan and G. Varghese, Fast address lookups using controlled prefix expansion, *ACM Trans. Comput. Syst.* **17** (1) (1999) 1–40.
31. W. Szpankowski, On a recurrence equation arising in the analysis of conflict resolution algorithms, *Comm. Statist. Stochastic Models* **3** (1) (1987) 89–114.
32. W. Szpankowski, Some results on V -ary asymmetric tries, *J. Algorithms* **9** (2) (1988) 224–244.
33. B.S. Tsybakov and V.A. Mikhaïlov, Free synchronous packet access in a broadcast channel with feedback, *Problems Inform. Transmission* **14** (4) (1978) 32–59.
34. N.D. Vvedenskaya and B.S. Tsybakov, Packet delay in a multiple-access stack algorithm, *Problemy Peredachi Informatsii* **20** (2) (1984) 77–97.