
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Paakkonen, Joonas; Dharmawansa, Prathapasinghe; Freij-Hollanti, Ragnar; Hollanti, Camilla; Tirkkonen, Olav

File Size Distributions and Caching for Offloading

Published in:

2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications, SPAWC 2019

DOI:

[10.1109/SPAWC.2019.8815572](https://doi.org/10.1109/SPAWC.2019.8815572)

Published: 01/07/2019

Document Version

Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:

Paakkonen, J., Dharmawansa, P., Freij-Hollanti, R., Hollanti, C., & Tirkkonen, O. (2019). File Size Distributions and Caching for Offloading. In *2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications, SPAWC 2019* Article 8815572 (IEEE International Workshop on Signal Processing Advances in Wireless Communications). IEEE. <https://doi.org/10.1109/SPAWC.2019.8815572>

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

File Size Distributions and Caching for Offloading

Joonas Pääkkönen*, Prathapasinghe Dharmawansa†, Ragnar Freij-Hollanti°, Camilla Hollanti°, Olav Tirkkonen*

* Department of Communications and Networking, Aalto University, Espoo, Finland

† Department of Electronic and Telecommunication Engineering, University of Moratuwa, Sri Lanka

° Department of Mathematics and Systems Analysis, Aalto University, Espoo, Finland

email: joonas.paakkonen@yahoo.com, prathapakd@gmail.com, ragnar.freij@aalto.fi,
camilla.hollanti@aalto.fi, olav.tirkkonen@aalto.fi

Abstract—We consider order statistics of products of random variables to understand the relationship between traffic offloading and storage space consumption in an edge caching network. Given a statistical model for the popularities and sizes of data objects in a caching network, we derive the order statistics describing the probability density function of the size of the object that accounts for, *e.g.*, the largest amount of data traffic. We further study the asymptotic behavior of these distributions, and give analytical results for both the general case and the asymptotic case. The derived results can be used to estimate an upper bound on the cache size needed when one is interested in offloading a certain fraction of all traffic for file catalogues with changing content but known distributions of popularities and object sizes.

Index Terms—Order Statistics, Caching

I. INTRODUCTION

Global mobile data traffic will increase sevenfold between 2017 and 2022 [1], passing 77 exabytes per month by 2022. To address this future capacity crunch, 5G technologies employ wider bandwidths, high-spectral efficiency massive-MIMO technologies, and ultra-dense networks [2]. Large bandwidths are available in mmWave frequencies, but the coverage of high-frequency networks is patchy, leading to potential disruptions in service.

The increase of memory capacity coupled with the decrease of memory prices leads to increased amounts of memory in mobile devices, and to the possibility to add memory to Radio Access Network (RAN) base stations, to be used for femto [3], D2D [4], [5] or edge caching [6]–[13].

We address proactive edge caching from the perspective of mobile network resource optimization. The underlying assumption is that in addition to an intermittently present high-capacity small-cell network, there is a macro-cellular network offering wide-area access. The macro-cellular network offers continuous coverage but moderate capacity, and is expected to be highly loaded. Proactive caching is performed at the user-end of the network, *e.g.*, at user devices, or in vehicles carrying the users and their equipment. The objective of network and cache management is to exploit the intermittent high-capacity

small-cell network in order to offload traffic from the macro-cellular network.

Several caching algorithms have been suggested in the literature, see *e.g.* [14] for a survey. Some of the most widely-used replacement algorithms are Least-Recently-Used (LRU), Least-Frequently-Used (LFU) and SIZE. When a cache object is about to be evicted, LRU/LFU evicts the least recently/frequently used objects, while SIZE evicts the largest objects. These methods are suitable for caching in the network, but for caching close to the users, such reactive methods may not apply, as an individual user, or the population of users in a small vehicle, may not consume the same content twice.

In this paper, we study a proactive edge caching scenario, where caching decisions are based on statistical inference. For this, we assume that statistics of file sizes and popularity are known. With a file cached at the user end, the expected amount of traffic offloaded from a macro network equals the product of the file size and popularity. This product we refer to as the *importance* of the file. Caching the most important file leads to the highest reduction in the amount of data traffic served by the macro-cellular network.

In this paper we address the question: What is the distribution of the size of the most important file, *i.e.* the file with the largest expected traffic, and more generally, of the k th least important file? To address this question, we derive the probability density function of the k th order statistic when considering a set of n samples of a variable which is a product of m random variables. We derive asymptotic bounds of this in the scaling limit $k/n \rightarrow \infty$. Concentrating on an example pertinent for caching at the user end, we provide a model for file size distribution of blue-ray movies, arguing that a gamma-distribution (as opposed to a fat-tailed distribution) well models the sizes of these large files. Then, assuming a Zipf popularity distribution, we consider the ensuing distribution of file importance, and compute the order statistics of this random variable both in the finite and asymptotic domain.

As an example related to offloading in an edge caching system, we consider the cumulative storage size needed to store a given number of most important files. This can be used to dimension cache size, given a target for expected traffic offloading.

This work was funded in part by TKK tukisäätiö and the Academy of Finland (AoF) grant #319058. The work of C. Hollanti was supported by AoF grants #276031, #282938, and #303819, and by the Technical University of Munich – Institute for Advanced Study, funded by the German Excellence Initiative and the EU 7th Framework Programme under grant agreement #291763, via a Hans Fischer Fellowship. The work of R. Freij-Hollanti is partially supported by the German Research Foundation grant WA3907/1-1.

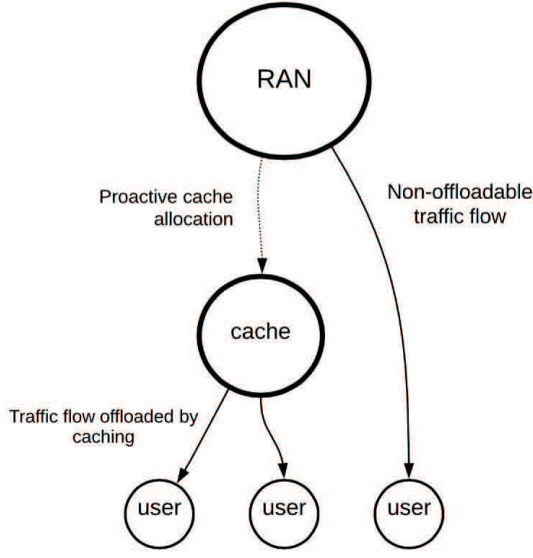


Fig. 1. Caching model. Files stored on the cache are served by the cache, while files not stored on the cache are retrieved from the umbrella network.

II. SYSTEM MODEL

Consider a network with users requesting files from an umbrella RAN with a cache between the users and the RAN according to Figure 1. If the file requested by a user is cached, the cache serves the user. Otherwise, the file is fetched from a content-delivery network, and the umbrella RAN transmits the file to the user. The cache contents are updated periodically, when the cache is at a high-bandwidth location in the radio access network. At each cache content update, the realizations of sizes and popularities of the cacheable files are sampled from given independent file size and popularity distributions, respectively. For each file, the importance metric is calculated as the product of file size and popularity. To find the required cache size for a given number of most important files, we need to find factor densities of order statistics as described in the following.

III. FACTOR DENSITIES OF ORDER STATISTICS

A. General Case

The k^{th} smallest random variate is commonly known as the k^{th} order statistic, where we emphasize that $k = 1$ refers to the smallest of the variates. We have n samples of a random variable U , which is a product of m random variables;

$$U = \prod_{i=1}^m X_i.$$

The factors of U thus are the independent random variables $\{X_i\}$, with densities $\{f_{X_i}\}$. Let $F_U(\cdot)$ denote the distribution function of U . We are interested in the density of an individual factor of the k^{th} smallest of the n samples of U .

Let \tilde{X} denote the factor of U in which we are interested after the ordering. In the following $f_X(\cdot)$ denotes the density

of the factor of interest before the ordering. Without loss of generality, we can let the factor of interest be the first factor of U . Further, let $\mathbb{E}_V\{\cdot\}$ denote the mathematical expectation with respect to the density of

$$V = \prod_{i=2}^m X_i,$$

and let us define

$$\mathcal{K}_{n,k} := \frac{n!}{(k-1)!(n-k)!}.$$

The factor density is given by the following result.

Theorem 1. *The probability density function of a factor of the k^{th} order statistic of a length- n sample of U is*

$$f_{\tilde{X}}(x) = \mathcal{K}_{n,k} f_X(x) \mathbb{E}_V\{F_U(xv)^{k-1} (1-F_U(xv))^{n-k}\}. \quad (1)$$

Proof. Let $f_B(k-1; n-1, p)$ be the probability mass function of the binomial distribution representing the event of $k-1$ successes out of $n-1$ trials with success probability p , defined as

$$f_B(k-1; n-1, p) := \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k}.$$

Without loss of generality, we can focus on the case where the k^{th} order statistic of a length- n sample of the product $U = \prod_{i=1}^m X_i$, denoted $U_{(k,n)}$, takes value u . By invoking the *mixed form of Bayes rule* (see, e.g., [15, 2.103a]), we obtain

$$f_{\tilde{X}}(x) := f_X(x|U_{(k,n)}=u) = \frac{\mathbb{P}(U_{(k,n)}=u|\tilde{X}=x) f_X(x)}{\mathbb{P}(U_{(k,n)}=u)}.$$

Finally, consider a given variate x , and a given variate t of $T = \prod_{i=2}^m X_i$. The probability that a product u is the k^{th} smallest product is equal to the probability that $k-1$ out of the remaining $n-1$ products take a value smaller than xt , which yields

$$\begin{aligned} & \mathbb{P}(U_{(k,n)}=u|\tilde{X}=x) \\ &= \int_{-\infty}^{\infty} f_B(k-1; n-1, F_U(xt)) dF_T \\ &= \mathbb{E}_T\{f_B(k-1; n-1, F_U(xt))\} \\ &= \binom{n-1}{k-1} \mathbb{E}_T\{F_U(xt)^{k-1} (1-F_U(xt))^{n-k}\}. \end{aligned}$$

The probability that an arbitrarily chosen u is the k^{th} smallest product out of n products is $\mathbb{P}(U_{(k,n)}=u) = 1/n$. Thus,

$$f_{\tilde{X}}(x) = \frac{\mathbb{E}_{x,T}\{f_B(k-1; n-1, F_U(xt))\} f_X(x)}{\mathbb{E}_U\{f_B(k-1; n-1, F_U(u))\}}$$

The statement follows directly. \square

B. Asymptotic Case

It appears difficult to characterize the impact of n and k on the factor distribution. To circumvent this difficulty, here we focus on the asymptotic regime, which allows us to find approximations for the density functions. In terms of the current caching application, this is especially important when the file catalogue size n is large.

We investigate the limit where $n \rightarrow \infty$ such that the ratio $k/n \rightarrow \eta$. This is referred to as the *double scaling limit*. Since the direct evaluation of such limits using (1) seems an arduous task, we adopt stochastic convergence concepts detailed in [16].

Let $g_m(\cdot)$ denote the density of $U = \prod_{i=1}^m X_i$, while $G_m(\cdot)$ denotes the cumulative distribution function of U and $G_m^{-1}(\cdot)$ denotes the inverse of $G_m(\cdot)$. Let $g_{m-1}(\cdot)$ denote the density of $V = \prod_{i=2}^m X_i$. Further, let $0 < \eta < 1$ and $\Phi = G_m^{-1}(\eta)$.

The asymptotic factor density is given by the following result.

Theorem 2. *As $n \rightarrow \infty$ such that $k/n \rightarrow \eta$, the asymptotic probability density function of a factor of the k^{th} order statistic of a length- n sample of U is*

$$f_{\bar{X}_a}(x) = \lim_{k/n=\eta, n \rightarrow \infty} f_{\bar{X}}(x) = \frac{g_{m-1}(\Phi/x)}{g_m(\Phi)|x|} f_X(x). \quad (2)$$

Proof. Let us rewrite (1) with a slightly modified notation as

$$f_{\bar{X}}(x) = \mathcal{K}_{n,k} f_X(x) \mathbb{E}_V \{ G_m(xv)^{k-1} (1 - G_m(xv))^{n-k} \}.$$

The expectation e over V in this expression can then be manipulated to yield

$$\begin{aligned} e &= \int_{-\infty}^{\infty} G_m^{k-1}(xv) (1 - G_m(xv))^{n-k} g_{m-1}(v) dv \\ &= \int_{-\infty}^{\infty} G_m^{k-1}(y) (1 - G_m(y))^{n-k} g_m(y) \frac{g_{m-1}(y/x)}{(xg_m(y))} dy, \end{aligned}$$

where the second equality follows from the variable transformation $xv = y$.

To proceed, let us use the variable transformation $z = G_m(y)$ with $dy = g_m(y)dy$ in order to arrive at another alternative form of (1) as

$$f_{\bar{X}}(x) = f_X(x) \int_0^1 g_{m-1}(G_m^{-1}(z)/x) / g_m(G_m^{-1}(z)) dp_{n,k}(z),$$

where

$$dp_{n,k}(z) = \Gamma(n+1) / (\Gamma(k)\Gamma(n-k+1)) z^{k-1} (1-z)^{n-k} dz$$

denotes a beta distribution with parameters k and $n-k+1$. Thus, the desired double scaling limit can be expressed as

$$\lim_{k/n \rightarrow \eta, n \rightarrow \infty} f_{\bar{X}}(x) = f_X(x) \int_0^1 \frac{g_{m-1}(y/x)}{g_m(y)|x|} dp_{n,k}(z).$$

The random variable Z_n associated with the beta density function $p_{n,\eta}(z)$ has the characteristic function

$$\begin{aligned} \phi_n(\omega) &= \int_0^1 \frac{\Gamma(n+1)}{\Gamma(k)\Gamma(n-k+1)} z^{k-1} (1-z)^{n-k} \exp(z\omega j) dz \\ &= {}_1F_1(k; n+1; -j\omega), \end{aligned}$$

where ${}_1F_1(\cdot; \cdot; \cdot)$ is the confluent hypergeometric function of the first kind and $j = \sqrt{-1}$.

Clearly, $\phi_n(\omega)$ converges $\forall \omega \in (-\infty, \infty)$ and $\phi_\infty(\omega) = \exp(j\omega\eta)$ and $\phi_\infty(\omega)$ is continuous at $\omega = 0$. Now note that the limiting characteristic function corresponds to a point mass at $z = \eta$. This alongside the fact that

$g_{m-1}(G_m^{-1}(z)/x) / g_m(G_m^{-1}(z))$ is bounded and continuous for $z \in (0, 1)$ enables us to invoke [16, Theorems 6.3.2 and 4.4.2] to yield

$$\begin{aligned} &\lim_{k/n \rightarrow \eta, n \rightarrow \infty} f_{\bar{X}}(x) \\ &= f_X(x) \lim_{n \rightarrow \infty} \int_0^1 \frac{g_{m-1}(G_m^{-1}(z)/x)}{g_m(G_m^{-1}(z))|x|} dp_{n,\eta}(z) \\ &= f_X(x) \int_0^1 \frac{g_{m-1}(G_m^{-1}(z)/x)}{g_m(G_m^{-1}(z))|x|} \delta(z - \eta) dz, \end{aligned}$$

where $\delta(z)$ is the Dirac delta function. Finally, using the sifting property of the delta function yields the desired result. \square

IV. SIZES OF THE IMPORTANT FILES

Here we derive the probability density function of the size of the k^{th} least important file when the number of files in the file catalogue is n for given size and popularity densities. This is done for the case with fixed k and n as well as the asymptotic case. Recall that the expected amount of data traffic that a file generates, *i.e.*, the importance of a file, is the product of its popularity and size.

A good model for the popularity of files is argued to be the so-called Zipf-law [17]. Then, the relative popularity of files in a catalogue follows the Zipf distribution with probability density function

$$f_P(p) = \frac{(pH_F^{(t)})^{-1/t}}{Ftp}, \quad (3)$$

where $H_F^{(t)} := \sum_{i=1}^F \frac{1}{i^t}$ is the harmonic number of order t and $p > \frac{F^{-t}}{H_F^{(t)}}$.

To find a model for the file size distribution, pertinent for a offloading scenario where videos are consumed in a vehicle, we study a large file repository [18], where the sizes of 5813 Blu-ray movies are shown in bytes. Based on these data, we find that the gamma distribution with parameters $\alpha = 8.978$, $\beta = 2.95 \cdot 10^9$ provides a good fit for the probability density function of the file size random variable S describing such a use case. The probability density function of the gamma distribution is

$$f_S(s) = \frac{\beta^{-\alpha} e^{-s/\beta} s^{\alpha-1}}{\Gamma(\alpha)} \quad (4)$$

with $s > 0$.

A straightforward calculation through (3) and (4) yields the density of the importance, *i.e.*, the density of the product of P and S , $U = PS$, as

$$f_U(u) = \frac{\beta^{-\alpha} u^{\alpha-1}}{\Gamma(\alpha)t} (\mathcal{I}(u) - \mathcal{J}(u)), \quad (5)$$

where

$$\mathcal{I}(x) := \Gamma(\alpha + 1/t) / \left(F q^{1/t} (x/\beta)^{\alpha+1/t} \right),$$

with

$$q := \zeta(t) - \zeta(t, F+1).$$

Here the gamma function is $\Gamma(\cdot)$, the Riemann zeta function is $\zeta(\cdot)$, and the Hurwitz zeta function is $\zeta(\cdot, \cdot)$. Furthermore

$$\mathcal{J}(x) := (F^t q)^\alpha E_R(\mathcal{Y}(x)),$$

with $\mathcal{Y}(x) := F^t q x / \beta$, $E(\cdot)$ the exponential integral function, and $R := -\alpha - 1/t + 1$. By integrating (5), the cumulative distribution function of U at $u = xv$ becomes

$$F_U(xv) = 1 - \frac{\Gamma(\alpha, \mathcal{Z})}{\Gamma(\alpha)} - \frac{\Gamma(\alpha + 1/t, \mathcal{Z}/q)}{\mathcal{Z}^{1/t} \Gamma(\alpha)} - \frac{\Gamma(\alpha + 1/t)}{\mathcal{Z}^{1/t} \Gamma(\alpha)}, \quad (6)$$

where $\mathcal{Z} := \mathcal{Z}(v, x) = v\mathcal{Y}(x)$ and $\Gamma(\cdot, \cdot)$ is the incomplete gamma function.

We now have

Corollary 1. *When file popularity follows Zipf(F, t) and file size follows Gamma(α, β), the probability density function of the size of the k^{th} least important file out of n files is*

$$f_{\bar{S}}(x) = \mathcal{K}_{n,k} \frac{\beta^{-\alpha} e^{-x/\beta} x^{\alpha-1}}{\Gamma(\alpha)} \int_{\frac{F-t}{H_F^{(t)}}}^{\infty} \left(1 - \frac{\Gamma(\alpha, \mathcal{Z})}{\Gamma(\alpha)} - \frac{\Gamma(\alpha + 1/t, \mathcal{Z}/q)}{\mathcal{Z}^{1/t} \Gamma(\alpha)} - \frac{\Gamma(\alpha + 1/t)}{\mathcal{Z}^{1/t} \Gamma(\alpha)} \right)^{k-1} \left(\frac{\Gamma(\alpha, \mathcal{Z})}{\Gamma(\alpha)} + \frac{\Gamma(\alpha + 1/t, \mathcal{Z}/q)}{\mathcal{Z}^{1/t} \Gamma(\alpha)} + \frac{\Gamma(\alpha + 1/t)}{\mathcal{Z}^{1/t} \Gamma(\alpha)} \right)^{n-k} \left(v H_F^{(t)} \right)^{-1/t} \frac{dv}{Ftv}. \quad (7)$$

Proof. Follows directly from using (3), (4) and (6) in (1) with $m = 2$. \square

Fig. 2 plots (7), with a file catalogue of $n = 10$ files, for $k \in \{1, 6, 10\}$. For comparison, the original parent gamma density (4), with $\alpha = 8.978$ and $\beta = 2.95 \cdot 10^9$, is plotted alongside the density of the maximum of 10 realizations of the file size random variable associated with the original parent density (4). For all numerical results in this section, we set $F = 5000$, $t = 0.6$ as in [19], $\alpha = 8.978$ and $\beta = 2.95 \cdot 10^9$.

Fig. 2 suggests key findings for a scenario with these distributions: We underline that the probability mass associated with the most important file is relatively close to the probability mass of the parent density. This in turn suggests that the most important files require only slightly more storage space compared to arbitrary files. Yet the most important files, by definition, offload most traffic.

We now turn our attention to the asymptotic behavior of the required cache size. To obtain the asymptotic factor density, we use $g_{m-1}(\Phi/x) = f_P(\Phi/x)$, $f_X(x) = f_S(x)$ and $g_m(\Phi) = f_U(\Phi)$ in (2). The asymptotic probability density function of the k^{th} least important file becomes

$$f_{\bar{S}_a}(x, \Phi) = \lim_{k/n \rightarrow \eta, n \rightarrow \infty} f_{\bar{S}}(x) = \frac{f_P(\Phi/x)}{f_U(\Phi)|x|} f_S(x), \quad (8)$$

where $\Phi = F_U^{-1}(\eta)$ is the inverse of the probability distribution function corresponding to the density of (5) at η .

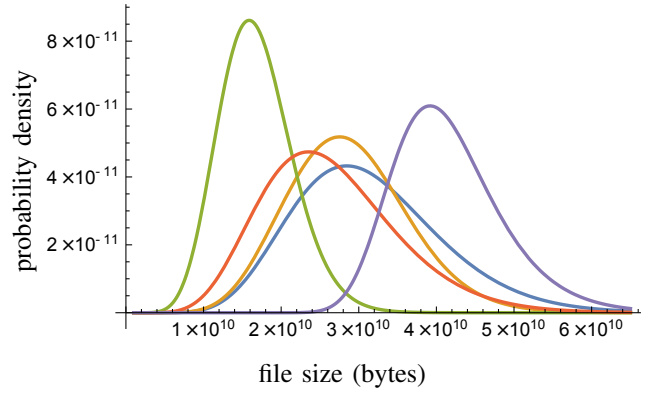


Fig. 2. File size densities (7) of the most important file ($k = n = 10$, blue), the fifth most important file ($k = 6$, orange) and the least important file ($k = 1$, green). The parent gamma density (4) is plotted in the red. The corresponding 10th order statistic is plotted in the purple.

We have

Corollary 2. *When file popularity follows Zipf(F, t) and file size follows Gamma(α, β), and as $n \rightarrow \infty$ such that $k/n \rightarrow \eta$, the asymptotic probability density function of the size of the k^{th} least important file out of n files is*

$$f_{\bar{S}_a}(x, \Phi) = x^{\alpha-1} e^{-x/\beta} \left(\frac{x}{\Phi H_F^{(t)}} \right)^{1/t} / \left(\frac{\Gamma(\alpha + 1/t) \beta^{\alpha+1/t}}{(\Phi q)^{1/t}} - (\Phi F^t q)^\alpha F E_R \left(\frac{\Phi F^t q}{\beta} \right) \right). \quad (9)$$

Proof. Follows from using (3), (4) and (5) in (8). \square

Lastly, we are interested in the expected cumulative cache size approximated by the asymptotic result (9) for large values of file catalogue size n , in the scenario with the considered distributions. The approximated expected size of the i^{th} most important file, with $i \in [2, n]$ and $\Phi = F_U^{-1}((n-i+1)/n)$, is

$$\mathcal{E}(i, n) = \int_{x=0}^{\Phi F^t H_F^{(t)}} x f_{\bar{S}_a}(x, \Phi) dx. \quad (10)$$

Fig. 3 shows the cumulative expected storage space consumption percentage of the most important files compared to the expected full file catalogue size when $n = 1000$. Numerical values are obtained through (10) for $i = 2 + 100\hat{i}$ with $\hat{i} \in [0, 9]$. Other values are obtained by interpolation, except for the approximated size of the most important file. For the most important file, with $i = 1$, extrapolation is applied.

The slight superlinearity of the blue curve in Fig. 3 suggests that the most important files require only slightly more storage space compared to the storage space requirements of an arbitrary cache allocation associated with the straight brown line.

V. CONCLUSIONS

We have derived probability density functions of factors of order statistics, and applied this theory to find the density

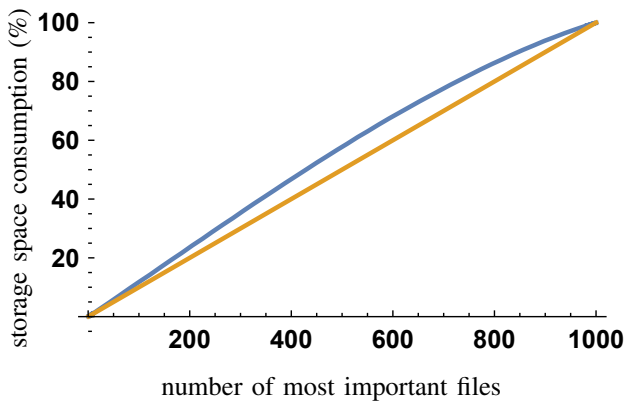


Fig. 3. Cumulative values of (10) (blue). A straight line (brown) is plotted for comparison. Here the catalogue size is $n = 1000$ files.

of the size of the file that in expectation generates the most data traffic. Furthermore, we have studied the corresponding asymptotic results. We have found that, for Gamma distributed file sizes, pertinent for a realistic video offloading scenario, and Zipf popularity models, the files that generate most traffic, *i.e.*, the most important files, require only moderately more storage space than an arbitrary file.

REFERENCES

- [1] Cisco, "Cisco visual networking index: Forecast and trends, 2017–2022 white paper," 2019.
- [2] A. Osseiran, J. Monserrat, and P. Marsch, eds., *5G Mobile and Wireless Communications Technology*. Cambridge University Press, 2016.
- [3] K. Shanmugam, N. Golrezaei, A. G. Dimakis, A. F. Molisch, and G. Caire, "Femto-caching: Wireless content delivery through distributed caching helpers," *IEEE Trans. Inf. Theory*, vol. 59, pp. 8402–8413, Dec. 2013.
- [4] N. Golrezaei, P. Mansourifard, A. F. Molisch, and A. G. Dimakis, "Base-station assisted device-to-device communications for high-throughput wireless video networks," *IEEE Trans. Wireless Commun.*, vol. 13, pp. 3665–3676, July 2014.
- [5] K. S. Khan, Y. Yin, and A. Jamalipour, "On the application of agglomerative hierarchical clustering for cache-assisted d2d networks," in *Proc. 16th IEEE Annual Consumer Commun. and Netw. Conf. (CCNC)*, Jan. 2019.
- [6] E. Baştuğ, M. Bennis, and M. Debbah, "Living on the edge: The role of proactive caching in 5g wireless networks," *IEEE Commun. Mag.*, vol. 52, pp. 82–89, Aug. 2014.
- [7] A. Liu and V. K. N. Lau, "How much cache is needed to achieve linear capacity scaling in backhaul-limited dense wireless networks?," *IEEE/ACM Trans. Netw.*, vol. 25, pp. 179–188, Feb. 2017.
- [8] D. Liu, B. Chen, C. Yang, and A. F. Molisch, "Caching at the wireless edge: design aspects, challenges, and future directions," *IEEE Commun. Mag.*, vol. 54, pp. 22–28, Sept. 2016.
- [9] S. Zhang, N. Zhang, P. Yang, and X. Shen, "Cost-effective cache deployment in mobile heterogeneous networks," *IEEE Trans. Veh. Technol.*, vol. 66, pp. 11264–11276, July 2017.
- [10] G. S. Paschos, G. Iosifidis, M. Tao, D. Towsley, and G. Caire, "The role of caching in future communication systems and networks," *IEEE J. Sel. Areas Commun.*, vol. 36, pp. 1111–1125, June 2018.
- [11] S. M. Azimi, O. Simeone, A. Sengupta, and R. Tandon, "Online edge caching and wireless delivery in fog-aided networks with dynamic content popularity," *IEEE J. Sel. Areas Commun.*, vol. 36, pp. 1189–1202, June 2018.
- [12] G. Vettigli, M. Ji, K. Shanmugam, J. Llorca, A. Tulino, and G. Caire, "Efficient algorithms for coded multicasting in heterogeneous caching networks," *Entropy*, vol. 21, Mar. 2019.
- [13] F. Lei, J. Cai, Q. Dai, and H. Zhao, "Deep learning based proactive caching for effective wsn-enabled vision applications," *Complexity*, May 2019.
- [14] S. Podlipnig and L. Böszörményi, "A survey of web cache replacement strategies," *ACM Comput. Surveys (CSUR)*, vol. 35, pp. 374–398, Dec. 2003.
- [15] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*. Wiley, 1965.
- [16] K. L. Chung, *A Course in Probability Theory*. Academic Press, 2001.
- [17] M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, and S. Moon, "I tube, you tube, everybody tubes: Analyzing the world's largest user generated content video system," in *Proc. 7th ACM SIGCOMM Conf. Internet Measurement (IMC)*, pp. 1–14, Oct. 2007.
- [18] AVSForum, "Unofficial blu-ray and video specifications thread." <http://www.avsforum.com/forum/150-blu-ray-software/1155731-new-unofficial-blu-ray-audio-video-specifications-thread.html>.
- [19] B. Chen and C. Yang, "Caching policy optimization for D2D communications by learning user preference," in *Proc. IEEE 85th Veh. Technol. Conf. (VTC Spring)*, June 2017.