Asymptotic Laws for Content Replication and Delivery in Wireless Networks

Savvas Gitzenis
Informatics & Telematics Institute
CERTH, Greece
E-mail: sgitz@iti.gr

Georgios S. Paschos
Informatics & Telematics Institute
CERTH, Greece
E-mail: gpasxos@iti.gr

Leandros Tassiulas
Dept. of Computer & Communication Eng.
University of Thessaly, Greece
E-mail:leandros@uth.gr

***Abstract*—A key consideration in novel communication paradigms in multihop wireless networks regards the scalability of the network. We investigate the case of nodes making random requests on content stored in multiple replicas over the wireless network. We show that, in contrast to the conventional paradigm of random communicating pairs, multihop communication is a sustainable scheme for certain values of file popularity, cache and network size. In particular, we formulate the joint problem of replication and routing and compute an order optimal solution. Assuming a Zipf file popularity distribution, we vary the number of files M in the system as a function of the nodes N, let both go to infinity and identify the scaling regimes of the required link capacity, from $O\left(\sqrt{N}\right)$ down to $O(1)$.**

I. Introduction

Networking based on content has been deemed as a key enabling technology for the future Internet [1]; routes are discovered and maintained based on the content request and provision. In this context, caching and prefetching information play a key role in various performance metrics (e.g., delay, throughput, scalability) and directly affect the Quality of Service (QoS) experienced by the user. Successful architectural paradigms like Content Delivery Networks (CDNs), Publish-Subscribe and Peer-to-Peer (P2P) systems underline the importance of caching as a means of preserving information over time and making it available in the vicinity of the user.

Wireless technology is another important enabler of the future Internet, as it promotes ubiquitous access for roaming users. Due to the volatile nature of wireless links and the associated performance bottlenecks, caching content becomes a challenging and important problem. Although it has been shown that wireless networks cannot sustain long multihop communications scenarios (the maximum common rate sustainable for all flows in the network scales inversely proportional to the number of hops [2]), it is unclear whether the addition of caching capability makes the system sustainable.

In this work, we study the impact of caching on the capacity of wireless networks, and make the following contributions: (i) we formulate the joint problem of replication and routing for minimizing the required wireless link capacity, and show that the solution is of the same order with the solution of the *Continuous Density* (CD) Problem, a mathematically tractable problem; (ii) we present the solution to the CD Problem, (iii) we investigate on the asymptotic laws on the link capacity and show that it ranges from $O(\sqrt{N})$, implying that the wireless network is not sustainable (as in [2]) down to $O(1)$, the latter implying that large enough wireless links suffice for the system to be sustainable; (iv) we provide the conditions between the number of files and the number of nodes so that caching makes a difference in the system performance; these conditions depend on the cache size and the file popularity distribution.

The paper is structured as follows: in Section II, we provide an overview of past work, and in Section III, we define the asymptotic notation. In Section IV, we define the **Worst Link** Problem, which considers the exact placement of data at each node along with their optimal delivery. This problem is of high complexity; therefore, to derive a computationally feasible solution, we present a series of related problems that culminate to the mathematically tractable **Continuous Density** Problem. The solution, presented in Section V, is within a constant factor of the original problem. Then, in Section VI, we identify on the asymptotic laws on the required link capacity when the number of nodes and files increase.

II. Problem Background & Related Work

Consider a wireless network of N nodes, where information is exchanged using multihop communications; the maximum throughput per node in such network scales as $O\left(\frac{1}{\sqrt{N}}\right)$ [2]. This celebrated, albeit pessimistic result essentially states that in large networks, per-user throughput is approximately zero. The degradation stems from the assumed uniform traffic matrix for the generation of throughput demand: the average communicating pair hop-distance increases as $\Theta\left(\sqrt{N}\right)$.

This work stimulated a series of attempts to breach the boundary, as in [3], where cooperation is used to mitigate the throughput drop (still hardly avoiding the $\sqrt{N}$ law). Nevertheless, the work in [4] established that the limitation is in nature geometrical, thus no physical model can breach this law given Maxwell's electromagnetic theory. In [5], non-uniform transfer matrices are studied leading to asymptotic laws for various types of flows (e.g., asymmetric, multicast, etc.). In [6], the case of a multihop wireless network aided by infrastructure is considered; to switch to a better law, a quite large number of base stations $\Omega\left(\sqrt{N}\right)$ is required.

In this study, we consider a modified delivery problem, with nodes generating requests on particular content (i.e., files), as opposed to particular destinations. Given the nodes' caching capability, files may exist in multiple nodes in the network; hence, requests are not necessarily directed to unique nodes. As usual, we assume a uniform distribution on the origin node of the requests, but a non-uniform distribution on the requested file. Then, we investigate if this delivery scenario is sustainable in a flat wireless network (i.e., overcome the result of [2]).

Caching has been proposed in the past as a means to improve wireless network performance. In [7], a caching scheme for wireless networks is proposed; performance benefits are shown thanks to minimization of wireless hops. In [8], caching is employed in the Publish-Subscribe paradigm to preserve information spatio-temporally and cover for link breakages and mobility. Cooperative caching for wireless meshes has been recently studied in [9] where it is demonstrated to improve the actual performance by means of implementation.

Here, we are interested in exploring the extent of the benefit of caching and whether its contribution to content delivery makes a difference in large networks. A similar study on scaling laws for caching in wireless networks can be found in [10]: the authors consider the case where an arbitrary traffic matrix is given and the information may be found in several caches in the network. Then, they determine the maximum rate of information that can be delivered to the destination. A series of interesting results follow; due to the fact that nodes must cooperatively guide the information to the destination, as in [3], the optimal strategy is not shortest path routing. Also, the network can become approximately sustainable in the same way with [3] using an hierarchical tree structure of transmissions over arbitrarily long links. However, it should be stressed that the results of [3], [10] depend on the particular signal attenuation parameters assumed.

In this study, we follow a different course from [3]: we fix a square grid as the wireless topology, a well known model for wireless networks [11]. In this flat topology, it remains unclear whether caching helps making the system scalable. Moreover, a second key difference in our approach is the fact that we consider the replication problem as in [7]: the positions of the files are part of the optimization problem, not pre-determined. In this context, we obtain similar results to [7].

Last, on the file request distribution, we focus on the Zipf law (as in [7]), as there exists ample evidence in the literature [12], [13] that the file popularity in the Internet follows such power laws. The Zipf parameter depends on the application, ranging from 0.5 [14] to 3 [15]; low values match the file distribution in routers, intermediate values in proxies and higher values in mobile applications [16], [17]; more studies exist in the literature and in the references of the above.

III. Asymptotic Notation

Let us define the notation used in the asymptotic laws that follow. Let f and g be real functions. Then, $f \in o(g)$ if

$$\text{for any } k > 0, \text{ there exists } x_\circ \text{ s. t. for } x \geq x_o, \left|\frac{f(x)}{g(x)}\right| \leq k.$$

Although $o(g)$ defines a set of functions, it is customary to write $f = o(g)$ (slightly abusing notation), instead of $f \in o(g)$.

Moreover, $f = O(g)$, if there exists a $k > 0$ such that $f(x)$ is eventually, in absolute value, less or equal to $kg(x)$, that is

$$\text{there exist } k > 0, x_o > 0 \text{ s. t. for } x \geq x_o, \left|\frac{f(x)}{g(x)}\right| \leq k.$$

Using such a k, we can write that $f \overset{\lim}{\leq} kg$ and $f \overset{\lim}{<} k'g$, if f, g are positive functions and $k' > k$.

Similarly, if the inequalities in the above definitions are reversed, i.e., $|f(x)/g(x)| \geq k$, then $f = \omega(g)$, or $f = \Omega(g)$, respectively. In the latter case, using such a k, we can write $f \overset{\lim}{\geq} kg$ and $f \overset{\lim}{>} k'g$, if f, g are positive functions and $k' < k$.

In the case that $f \overset{\lim}{\leq} g$ and $f \overset{\lim}{\geq} g$, we write that $f \sim g$.

Last, $f = \Theta(g)$ if $f = \Omega(g)$ and $f = O(g)$.

An important consequence of the above is that $f = O(g)$ does not imply $f \overset{\lim}{<} g$—e.g., consider $f(x) = 2g(x)$; however, the reverse is true. Moreover, if $f \overset{\lim}{<} g$, then $g - f = \Theta(g)$.

IV. The Replication Problem

Consider N nodes with N being a square of an integer, arranged on a square lattice on the plane of size $\sqrt{N}$ rows times $\sqrt{N}$ columns. Each node is connected to its four immediate neighbors on the same row or column, creating a flat grid topology. Keeping the node density fixed and increasing N, the network scales as in [2]. Moreover, to avoid boundary effects, we consider a toroidal structure [18].

This grid structure permits to consider the discrete nature of the wireless nodes, unlike the average node/cache capacity density of [7]. Unless nodes cooperate in their transmissions in a complex scheme (e.g., [3], [10]), it is a reasonable topology to consider, as wireless nodes have a limited communication range due to path attenuation and interference. With a frequency reuse factor appropriate to the physical layer, and/or a TDMA scheme or a random access scheme at the MAC layer, we can abstract away the network layer view to the lattice graph, as in [11]. Moreover, this is a setup that essentially captures the result of [2] for random communicating pairs.

Nodes (or users at the nodes) place requests on files/data, indexed by $m \in \mathcal{M} \triangleq \{1, 2, \ldots, M\}$. Each node is equipped with a buffer/cache. A request generated on a node about a file m stored in the node's cache is directly served, without using the network. Otherwise, the request is served from another node that keeps m in its cache, generating network traffic. Nodes are assumed alike, each having a cache of capacity K, i.e., can store K files. This implies that all files are of equal size (big files can be modeled as multiple indepedent unit-size files). For the problem to have a solution, it should be

$$KN \geq M, \tag{1}$$

otherwise the network has insufficient memory to store each file at least once. Moreover, for the problem of replication not to be trivial, it should be $K < M$: each node must make a specific selection of files to buffer in its cache.

Let each node $n \in \mathcal{N}$ generate requests for files at a total common rate of λ. Each request regards a particular data $m \in \mathcal{M}$, depending on the file's *popularity* p_m. In essence, $[p_m]$ is a probability distribution that dictates replication: to minimize the network traffic, popular data should be stored densely.

From the Microscopic to Macroscopic View

In the above setup, we are interested in optimizating the network operations. The subject is twofold and regards

- $\mathcal{B}$: the replication policy, which specifies the buffer (cache) contents across all nodes, and
- $\mathcal{R}$: the routing policy, which specifies how requests of any file m are routed in the network from any node n to a node n' that can serve the request for file m.

In our study, we estimate the rate $C_\ell(\mathcal{B},\mathcal{R})$ on link ℓ given the replication and routing policies $\mathcal{B}$ and $\mathcal{R}$. This rate sets the minimum capacity of link ℓ required to sustain the resulting traffic. Clearly, link rates C_ℓ are proportional to the request rate λ. Hence, without loss of generality, in the rest of this work, we assume that $\lambda = 1$. In contrast, in [2], a constant link rate is assumed (e.g. $C_\ell = 1$) and then, the maximum λ is computed; this is essentially the inverse perspective.

The objective of the optimization problem is to find the *joint* replication and routing policies $\mathcal{B}$ and $\mathcal{R}$ that minimize the required link capacities $C_\ell(\mathcal{B},\mathcal{R})$. In particular, in the primary formulation of the problem, we would like to consider the *worst* case, that is the rate of the most loaded link:

PROBLEM 1 [WORST LINK]: $C_{\mathtt{WL}} = \min_{\mathcal{B},\mathcal{R}} \max_{\ell} \; C_\ell(\mathcal{B},\mathcal{R})$.

Clearly, this problem is of combinatorial complexity, and thus not amenable to an easy to compute solution. However, as our focus is on the asymptotics, we state a set of gradually simplifying problems; these lead to a straighforward replication strategy whose solution is within a constant to the WL problem. Due to space limitations, this section provides a high-level overview only (proofs can be found in [19]).

A first step is the Average Link Problem, focusing on the average rate across network, as opposed to the worst case:

PROBLEM 2 [AVERAGE LINK]: $C_{\mathtt{AL}} = \min_{\mathcal{B},\mathcal{R}} \operatorname{avg}_{\ell} C_\ell(\mathcal{B},\mathcal{R})$.

From the definition of the problems, it is clear that $C_{\mathtt{AL}} \leq C_{\mathtt{WL}}$. Moreover, AL is an easier problem than WL (albeit still of combinatorial complexity due to the selection of cache contents at each node): shortest path routing to the nearest node that holds file m suffices in AL (proof in [19]).

To further simplify the problem, we dispense with the N individual buffer constraints at each node, and specify instead a *total cache capacity* KN across the network. We can then create new replication policies $\overline{\mathcal{B}}$ with more than K files in some nodes, and less than K files in other nodes. Defining

PROBLEM 3 [TOTAL CACHE (TC) CONSTRAINT]:
$$C_{\mathtt{TC}} = \min_{\overline{\mathcal{B}},\mathcal{R}} \operatorname{avg}_{\ell} C_\ell(\overline{\mathcal{B}},\mathcal{R}),$$

it is $C_{\mathtt{TC}} \leq C_{\mathtt{AL}}$, as any solution of AL satisfies TC, as well.

In the above problems, there is a 'microscopic' decision to focus on the cache contents of every node. To eliminate the combinatorial buffer configurations, we switch our attention to the frequency of occurrence of each file m in the caches, and define a macroscopic metric, the *replication density* d_m as the fraction of nodes that store file m. The inverse d_m^{-1} corresponds, in a fluid approximation, to the number of nodes served by a node that maintains m in its cache [7], or the size of the area served by this node. Moreover, under shortest path routing, the number of hops required from a node n to reach a cache containing file m can be shown to be on average $k_{\mathtt{CD}}\left(\frac{1}{\sqrt{d_m}} - 1\right)$, with $k_{\mathtt{CD}} = \sqrt{2}/6$ for the lattice topology.

Note that in all previous problems, d_m is a number k/N, with k a positive integer up to N, and moreover $\sum_{m\in\mathcal{M}} d_m \leq K$. We can then define an abstract problem regarding the densities of the data m instead of the cache contents:

PROBLEM 4 [CONTINUOUS DENSITY]:
$$C_{\mathtt{CD}} = \min_{[d_m]} \sum_{m\in\mathcal{M}} k_{\mathtt{CD}}\left(\frac{1}{\sqrt{d_m}} - 1\right) p_m, \text{ subject to:}$$

1) *For any* $m \in \mathcal{M}$, $\frac{1}{N} \leq d_m \leq 1$,
2) $\sum_{m\in\mathcal{M}} d_m \leq K$.

Any solution of the previous problems results in a density vector $[d_m]$ that satisfies CD Problem constraints. Therefore,

$$C_{\mathtt{CD}} \leq C_{\mathtt{TC}} \leq C_{\mathtt{AL}} \leq C_{\mathtt{WL}}.$$

In other words, the CD problem establishes a lower bound on the WL. Moreover, there exist positive $k_{\mathtt{WL}}$ and $k'_{\mathtt{WL}}$, such that we can always map a density vector $[d_m]$ satisfying the CD problem to a (not necessarily optimal) buffer and routing policy of the WL, whose worst link rate is up to $k_{\mathtt{WL}}C_{\mathtt{CD}} + k'_{\mathtt{WL}}$ (proof in [19]). Thus, $C_{\mathtt{WL}} \leq k_{\mathtt{WL}}C_{\mathtt{CD}} + k'_{\mathtt{WL}}$. Although $k'_{\mathtt{WL}}$ is a function of M, overall $C_{\mathtt{WL}} = O(C_{\mathtt{CD}})$, thus, for the asymptotic laws, it suffices to consider the CD problem only.

Note that [7] studies a problem of the same optimization target in a more relaxed topology than the lattice, without considering the discrete nodes and the associated constraints on the values of each d_m; as seen next, these play a major role in the asymptotics of the various τ and M vs. N regimes.

V. SOLUTION OF THE REPLICATION PROBLEM

The CD Problem is a monotropic optimization problem with a nonlinear objective function and linear constraints. To find the solution, we have to consider $2M+1$ Lagrange multipliers, and carry out a search on the binding constraints. The result partitions the set of files into three subsets (proof in [19]):

- $\mathcal{M}_{\uparrow} = \{1, 2, \ldots, l-1\}$ contains files of unit replication density $d_m = 1$ (i.e., to be stored to every node)
- $\mathcal{M}_{\updownarrow} = \{l, l+1, \ldots, r-1\}$ contains files stored multiple times, but not everywhere ($\frac{1}{N} < d_m < 1$),
- $\mathcal{M}_{\downarrow} = \{r, r+1, \ldots, M\}$ contains files stored once, $d_m = \frac{1}{N}$,

where l and r index the first file of $\mathcal{M}_{\updownarrow}$ and $\mathcal{M}_{\downarrow}$, respectively, for the files ordered in decreasing popularity, as in (3), as illustrated in Fig. 1. Then, the densities of the files are

$$d_m = \begin{cases} 1, & m \in \mathcal{M}_{\uparrow}, \quad \text{(2a)} \\ \left(K - l + 1 - \dfrac{M - r + 1}{N}\right) \dfrac{p_m^{\frac{2}{3}}}{\sum_{j=l}^{r-1} p_j^{\frac{2}{3}}}, & m \in \mathcal{M}_{\updownarrow}, \quad \text{(2b)} \\ \dfrac{1}{N}, & m \in \mathcal{M}_{\downarrow}. \quad \text{(2c)} \end{cases}$$

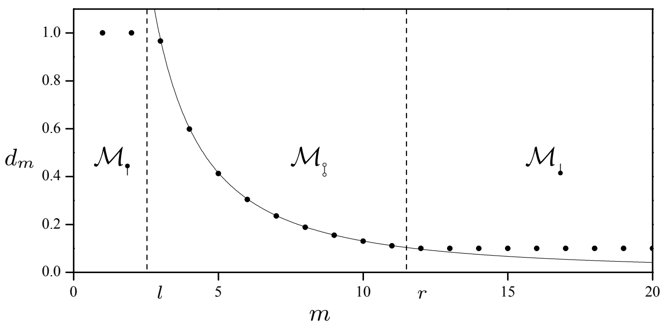

Fig. 1. An example case of density d_m and the $\mathcal{M}_{\uparrow}, \mathcal{M}_{\updownarrow}$ and $\mathcal{M}_{\downarrow}$ partitions.

The following result (see [19] for its proof) establishes the uniqueness of indices l, r and sets $\mathcal{M}_{\uparrow}, \mathcal{M}_{\updownarrow}, \mathcal{M}_{\downarrow}$:

Theorem 1 [Uniqueness]: *The indices (l, r) of the optimal solution of the CD problem are unique. Moreover, it is not possible to decrease l or increase r given the optimal form (2) without violating the constraints $\frac{1}{N} \le d_m \le 1$.*

A. Zipf Law and Approximations

We consider the Zipf law for file popularity as follows:

$$p_m = \frac{m^{-\tau}}{\sum_{m=1}^{M} m^{-\tau}} = \frac{m^{-\tau}}{H_\tau(M)}, \tag{3}$$

where τ is the power law parameter, indicating the rate of popularity decline as m increases. Moreover, $H_\tau(n) \triangleq \sum_{j=1}^{n} j^{-\tau}$ is the truncated (at n) zeta function evaluated at τ (also called the n^{th} *τ-order generalized harmonic number*). The limit $H_\tau \triangleq \lim_{n\to\infty} H_\tau(n)$ is the Riemann zeta function, which converges when $\tau > 1$. We derive an approximation for $H_\tau(n)$ by bounding the sum by two integrals; for $n \ge m \ge 0$, it is

$$\int_m^n (x+1)^{-\tau} dx \le H_\tau(n) - H_\tau(m) \le 1 + \int_{m+1}^n x^{-\tau} dx, \Rightarrow$$
$$\begin{cases} \frac{(n+1)^{1-\tau} - (m+1)^{1-\tau}}{1-\tau} \le H_\tau(n) - H_\tau(m) \le \frac{n^{1-\tau} - (m+1)^{1-\tau}}{1-\tau} + 1, & \text{if } \tau \ne 1, \\ \ln \frac{n+1}{m+1} \le H_\tau(n) - H_\tau(m) \le \ln \frac{n+1}{m+2}, & \text{if } \tau = 1. \end{cases} \tag{4}$$

B. Basic Properties of the Solution

As we are interested in the asymptotic scaling of link rates, for notational simplicity, we remove the k_{CD} factor and refer to the resulting quantity as C. Substituting the solution (2) and the Zipf distribution (3) into the objective function,

$$C = \sum_{m\in\mathcal{M}} \left(\frac{1}{\sqrt{d_m}} - 1 \right) p_m = C_{\updownarrow} + C_{\downarrow} - \sum_{j=l}^{M} p_m, \tag{5}$$

where

$$C_{\updownarrow} \triangleq \sum_{m\in\mathcal{M}_{\updownarrow}} \frac{p_m}{\sqrt{d_m}} \overset{(3)}{=} \frac{\left[H_{\frac{2\tau}{3}}(r-1) - H_{\frac{2\tau}{3}}(l-1) \right]^{\frac{3}{2}}}{\sqrt{K_{\updownarrow}}\, H_\tau(M)}, \tag{6}$$

$$C_{\downarrow} \triangleq \sum_{m\in\mathcal{M}_{\downarrow}} \frac{p_m}{\sqrt{d_m}} \overset{(3)}{=} \sqrt{N}\, \frac{H_\tau(M) - H_\tau(r-1)}{H_\tau(M)}, \tag{7}$$

$$K_{\updownarrow} \triangleq \frac{(K-l+1)N - (M-r+1)}{N}. \tag{8}$$

Note that $\sum_{j=l}^{M} p_m = O(1)$, as it lies always in $[0, 1]$.

C. Estimation of l and r

As indices l, r are not provided in a closed form, we have to derive suitable approximations to study the scaling of C.

1) Estimation of l: note that $l \le K+1$ (hence $l = \Theta(1)$), as $l-1$ is the number of files replicated in all nodes. If $\mathcal{M}_{\updownarrow}$ and $\mathcal{M}_{\downarrow}$ are not empty, using (2b), $d_l < 1$ is equivalent to

$$K - l + 1 - \frac{M-r+1}{N} < l^{\frac{2\tau}{3}} \left[H_{\frac{2\tau}{3}}(r-1) - H_{\frac{2\tau}{3}}(l-1) \right]. \tag{9}$$

If, moreover, the first set $\mathcal{M}_{\downarrow}$ is not empty, i.e., $l > 1$, then $d_{l-1} = 1$. This means that if we attempted to decrease index l by 1, this would violate the density constraints (Theorem 1), and get from (2b) a number greater than 1 for d_{l-1}:

$$K - l - \frac{M-r+1}{N} \ge (l-1)^{\frac{2\tau}{3}} \left[H_{\frac{2\tau}{3}}(r-1) - H_{\frac{2\tau}{3}}(l-2) \right]. \tag{10}$$

Thus, provided $l > 1$, it can be uniquely determined as the lowest integer that satififies (9-10), which is unique (Theorem 1). An approximation for l can be computed treating (9) as an approximate equality when $\mathcal{M}_{\updownarrow} \ne \emptyset$, or equivalently when $l < r$ (as $d_{l-1} = 1$ and $d_l < 1$):

$$K - l + 1 - \frac{M-r+1}{N} \cong l^{\frac{2\tau}{3}} \left[H_{\frac{2\tau}{3}}(r-1) - H_{\frac{2\tau}{3}}(l-1) \right]. \tag{11}$$

2) Estimation of r: If $\mathcal{M}_{\updownarrow} \cup \mathcal{M}_{\uparrow}$ is not empty, $d_{r-1} > \frac{1}{N} \Leftrightarrow$

$$(K-l+1)N - M + r - 1 > (r-1)^{\frac{2\tau}{3}} \left[H_{\frac{2\tau}{3}}(r-1) - H_{\frac{2\tau}{3}}(l-1) \right]. \tag{12}$$

Again, if the third set $\mathcal{M}_{\downarrow}$ is not empty, i.e., $r \le M$, it is $d_r = 1/N$. Thus, if we attempted increasing index r by one, (2b) would yield a density less than $1/N$ (Theorem 1):

$$(K-l+1)N - M + r \le r^{\frac{2\tau}{3}} \left[H_{\frac{2\tau}{3}}(r) - H_{\frac{2\tau}{3}}(l-1) \right]. \tag{13}$$

As before, (12) is an approximate equality if $l < r$, i.e.,

$$(K-l+1)N - M + r - 1 \cong (r-1)^{\frac{2\tau}{3}} \left[H_{\frac{2\tau}{3}}(r-1) - H_{\frac{2\tau}{3}}(l-1) \right]. \tag{14}$$

3) Estimation of l/r: For all l, r, it is $N > \frac{d_l}{d_{r-1}} = \left(\frac{r-1}{l} \right)^{\frac{2\tau}{3}}$. In the spirit of the previous analyses, whenever both l and r are not equal to the extremes, i.e., $1 < l < r < M+1$, the ratio $d_{l-1}/d_r = N$ holds also true. Thus,

$$l \cong \frac{r}{N^{\frac{3}{2\tau}}}. \tag{15}$$

The proof of all the following results can be found in the Appendix. The following Lemma shows that it is impossible to have files cached in *all* nodes unless the file popularity parameter τ is greater than $\frac{3}{2}$:

Lemma 2: *If $\tau \le \frac{3}{2}$, then $l \to 1$.*

VI. Asymptotic Laws

We proceed in studying the asymptotic behavior of the system regarding the link rate C, as well as indices l and r, which govern the size of partitions $\mathcal{M}_\uparrow, \mathcal{M}_\updownarrow$ and $\mathcal{M}_\downarrow$. The derived asymptotic laws regard the scaling when the number of nodes N and the number of files M increase to infinity. We use $\hat{l}$ and $\hat{r}$ to refer to the limits of l and r.

First, we provide a set of basic results that establish the upper bound of C equal to $O\left(\sqrt{N}\right)$. This is the Gupta-Kumar rate [2], a result intuitively expected: if replication is ineffective (e.g., due to large number of files), then the system essentially reduces to [2], matching its performance.

Lemma 3 [Bound on $C_\updownarrow$]: $C_\updownarrow = O\left(\sqrt{N}\right)$.

Lemma 4 [Bounds on $C_\downarrow$]: $C_\downarrow = O\left(\sqrt{N}\right)$. *Furthermore,*

1) *for* $\tau < 1$, *and* $r \overset{\lim}{<} M$, *it is* $C_\downarrow = \Theta\left(\sqrt{N}\right)$,
2) *for* $\tau > 1$, *it is* $C_\downarrow = \Theta\left(\sqrt{N}\left(r^{1-\tau} - M^{1-\tau}\right)\right)$.
   If, moreover, $r \overset{\lim}{<} M$, *then* $C_\downarrow = \Theta\left(\frac{\sqrt{N}}{r^{\tau-1}}\right)$.

Corollary 5 [Bound on C]: $C = O\left(\sqrt{N}\right)$.

Next, we start the analysis by partitioning the space of M, N on whether they produce single replicated files or not.

A. Almost Empty $\mathcal{M}_\downarrow$

The first case to consider is when the number of nodes N and files M increase towards infinity, and at the same time $\mathcal{M}_\downarrow$ remains an *almost empty* set. We define formally $\mathcal{M}_\downarrow$ to be almost empty, through $|\mathcal{M}_\downarrow| = o(M)$, i.e., the number of its elements is of lower order than the total files. For this to happen, M should increase at a slow pace with N, so that the constraint $d_m \geq 1/N$ is satisfied for almost all files, or equivalently $d_m = 1/N$ for $o(M)$ files. The extreme case of this is to first let $N \to \infty$ and then $M \to \infty$, i.e., split the limit of N, M jointly going to infinity to a *double* limit.

To study the asymptotics of C, we first estimate l and r. The almost empty $\mathcal{M}_\downarrow$ implies that $|\mathcal{M}_\downarrow| = M - r = o(M)$, and thus $r \sim M$ (i.e., as $r/M \to 1$).

Theorem 6 [$\hat{l}$ for Almost Empty $\mathcal{M}_\downarrow$]:

1) *For* $\tau \leq \frac{3}{2}$, $l \to 1$.
2) *For* $\tau > \frac{3}{2}$, $l \to \hat{l}_{\{\tau>\frac{3}{2}, \mathcal{M}_\downarrow \approx \emptyset\}}$, *where* $\hat{l}_{\{\tau>\frac{3}{2}, \mathcal{M}_\downarrow \approx \emptyset\}}$ *is the integer solution of*

$$\begin{cases} (K-l+1)l^{-\frac{2\tau}{3}} & < \quad H_{\frac{2\tau}{3}} - H_{\frac{2\tau}{3}}(l-1), \\ (K-l)(l-1)^{-\frac{2\tau}{3}} & \geq \quad H_{\frac{2\tau}{3}} - H_{\frac{2\tau}{3}}(l-2), \end{cases} \tag{16}$$

if such exists and is greater than 1, or 1 otherwise.

An approximation of (16) for $K \gg 1$ is the following:

$$K - (l-1) \cong (l-1)\left[H_{\frac{2\tau}{3}} - H_{\frac{2\tau}{3}}(l-1)\right] \overset{(4)}{\cong} \frac{l-1}{\frac{2\tau}{3}-1} \Leftrightarrow$$

$$l \cong 1 + \frac{2\tau - 3}{2\tau} K. \tag{17}$$

Next, we find the conditions so that $\mathcal{M}_\downarrow$ is almost empty:

Theorem 7 [$\mathcal{M}_\downarrow$ Almost Empty]: *It is* $M - r = o(M)$ *iff*

$$\begin{cases} M \overset{\lim}{\leq} \left(1 - \frac{2\tau}{3}\right) KN, & \text{if } \tau < \frac{3}{2}, \\ M \ln M \overset{\lim}{\leq} KN, & \text{if } \tau = \frac{3}{2}, \\ M \overset{\lim}{\leq} \left[\frac{(K-\hat{l}+1)\left(\frac{2\tau}{3}-1\right)}{\hat{l}^{1-\frac{2\tau}{3}}}\right]^{\frac{3}{2\tau}} N^{\frac{3}{2\tau}}, & \text{if } \tau > \frac{3}{2}. \end{cases}$$

where $\hat{l} = \hat{l}_{\{\tau>\frac{3}{2}, \mathcal{M}_\downarrow = \emptyset\}}$ *from Theorem 6. If the above inequalities are strict, then* $r = M + 1$ *(and thus* $\mathcal{M}_\downarrow = \emptyset$*).*

Theorem 8 [Capacity for Almost Empty $\mathcal{M}_\downarrow$]:

$$C = \begin{cases} \Theta\left(\sqrt{M}\right), & \text{if } \tau < 1, \\ \Theta\left(\frac{\sqrt{M}}{\log M}\right), & \text{if } \tau = 1, \\ \Theta\left(M^{\frac{3}{2}-\tau}\right), & \text{if } 1 < \tau < \frac{3}{2}, \\ \Theta\left((\log M)^{\frac{3}{2}}\right), & \text{if } \tau = \frac{3}{2}, \\ \Theta(1), & \text{if } \tau > \frac{3}{2}. \end{cases}$$

B. Non-empty $\mathcal{M}_\downarrow$

When $M_\downarrow$ is non-empty, it is $C_\downarrow > 0$. As Corollary 5 shows, C is $O\left(\sqrt{N}\right)$, i.e., the Gupta-Kumar rate [2]. Thus, we turn our attention to identifying the cases that $C = o\left(\sqrt{N}\right)$.

Theorem 9 [$\hat{l}$ and $\hat{r}$ for Non-Empty $\mathcal{M}_\downarrow$]: *If* M *exceeds the condition of Theorem 6,*

- *if* $KN - M = \omega(1)$, *then we discern the following cases:*

$$\tau < \frac{3}{2}: \quad l \to 1, \quad r \sim \frac{3-2\tau}{2\tau}(KN - M), \tag{18}$$

$$\tau = \frac{3}{2}: \quad l \to 1, \quad r \ln r \sim KN - M, \tag{19}$$

$$\tau > \frac{3}{2} \text{ and } M \overset{\lim}{\leq} (K - \beta) N:$$

$$l \to \hat{l} \cong \alpha\left[K + 1 - \lim \frac{M}{N}\right], \tag{20}$$

$$r \sim \alpha\left[KN^{\frac{3}{2\tau}} - \frac{M}{N^{1-\frac{3}{2\tau}}}\right]. \tag{21}$$

$$\tau > \frac{3}{2} \text{ and } M \overset{\lim}{>} (K - \beta) N:$$

$$l \to 1, \quad r \sim \left[\frac{2\tau}{3}(KN - M)\right]^{\frac{3}{2\tau}} \tag{22}$$

  where $\alpha = \frac{2\tau-3}{2\tau}$, $\beta = \frac{3}{2\tau-3}$.
- *if* $KN - M = O(1)$, *then* $l \to 1, r = \Theta(1)$, *with the exact value determined by*

$$\begin{cases} KN - M + r - 1 > (r-1)^{\frac{2\tau}{3}} H_{\frac{2\tau}{3}}(r-1), \\ KN - M + r \leq r^{\frac{2\tau}{3}} H_{\frac{2\tau}{3}}(r). \end{cases} \tag{23}$$

Note that the approximation of (20) on $\hat{l}$ can be precisely carried out via (16), if we substitute K with $K - \lim M/N$.

TABLE I

(a) The Cases of $\tau < 1$, $\tau = 1$ and $1 < \tau < \frac{3}{2}$.

M		M finite	$N \to \infty$ then $M \to \infty$	$M \overset{\lim}{\leq} \frac{3-2\tau}{2\tau} KN$	$M \overset{\lim}{>} \frac{3-2\tau}{2\tau} KN$ and $M \overset{\lim}{<} KN$	$M \sim KN$	
						$KN - M = \omega(1)$	$KN - M = O(1)$
$\mathcal{M}_\downarrow$		empty	empty	almost empty	non-empty	non-empty	non-empty
$\hat{l}$		1	1	1	1	1	1
$\hat{r}$		$M+1$	$M+1$	$M - o(M)$	$\frac{3-2\tau}{2\tau}(KN - M)$	$\frac{3-2\tau}{2\tau}(KN - M)$	$\Theta(1)$ (23)
C	$\tau < 1$	$\Theta(1)$	$\Theta\left(\sqrt{M}\right)$	$\Theta\left(\sqrt{M}\right)$	$\Theta\left(\sqrt{M}\right)$	$\Theta\left(\sqrt{M}\right)$	$\Theta\left(\sqrt{M}\right)$
	$\tau = 1$	$\Theta(1)$	$\Theta\left(\frac{\sqrt{M}}{\log M}\right)$	$\Theta\left(\frac{\sqrt{M}}{\log M}\right)$	$\Theta\left(\frac{\sqrt{M}}{\log M}\right)$	$\Theta\left(\sqrt{M}\right)$	$\Theta\left(\sqrt{M}\right)$
	$1 < \tau < \frac{3}{2}$	$\Theta(1)$	$\Theta\left(M^{\frac{3}{2}-\tau}\right)$	$\Theta\left(M^{\frac{3}{2}-\tau}\right)$	$\Theta\left(M^{\frac{3}{2}-\tau}\right)$	$\Theta\left(\frac{\sqrt{M}}{(KN-M)^{\tau-1}}\right)$	$\Theta\left(\sqrt{M}\right)$

(b) The Case of $\tau = \frac{3}{2}$.

M	M finite	$N \to \infty$ then $M \to \infty$	$M \ln M \overset{\lim}{\leq} KN$	$M \ln M \overset{\lim}{>} KN$ and $M \overset{\lim}{<} KN$	$M \sim KN$	
					$KN - M = \omega(1)$	$KN - M = O(1)$
$\mathcal{M}_\downarrow$	empty	empty	almost empty	non-empty	non-empty	non-empty
$\hat{l}$	1	1	1	1	1	1
$\hat{r}$	$M+1$	$M+1$	$M - o(M)$	$r \ln r \sim KN - M$	$r \ln r \sim KN - M$	$\Theta(1)$ (23)
C	$\Theta(1)$	$\Theta\left((\log M)^{\frac{3}{2}}\right)$	$\Theta\left((\log M)^{\frac{3}{2}}\right)$	$\Theta\left((\log r)^{\frac{3}{2}}\right)$	$\Theta\left(\sqrt{\frac{M}{KN-M}}(\log r)^{\frac{3}{2}}\right)$	$\Theta\left(\sqrt{M}\right)$

(c) The Case of $\tau > \frac{3}{2}$.

M	M finite	$N \to \infty$ then $M \to \infty$	$M \overset{\lim}{\leq} hN^{\frac{3}{2\tau}}$ (see Theorem 7)	$M \overset{\lim}{>} hN^{\frac{3}{2\tau}}$ and $M \overset{\lim}{\leq} (K-\beta)N$	$M \overset{\lim}{>} (K-\beta)N$ and $M \overset{\lim}{<} KN$	$M \sim KN$	
						$KN - M = \omega(1)$	$KN - M = O(1)$
$\mathcal{M}_\downarrow$	empty	empty	almost empty	non-empty	non-empty	non-empty	non-empty
$\hat{l}$	$\Theta(1)$ (16)	$\Theta(1)$ (16)	$\Theta(1)$ (16)	$\cong \alpha\left[K+1-\lim \frac{M}{N}\right]$	1	1	1
$\hat{r}$	$M+1$	$M+1$	$M - o(M)$	$\sim\alpha\left[KN^{\frac{3}{2\tau}} - \frac{M}{N^{1-\frac{3}{2\tau}}}\right]$	$\sim\left[\frac{2\tau}{3}(KN-M)\right]^{\frac{3}{2\tau}}$	$\sim\left[\frac{2\tau}{3}(KN-M)\right]^{\frac{3}{2\tau}}$	$\Theta(1)$ (23)
C	$\Theta(1)$	$\Theta(1)$	$\Theta(1)$	$\Theta(1)$	$\Theta(1)$	$\Theta\left(\frac{\sqrt{M}}{(KN-M)^{\frac{3(\tau-1)}{2\tau}}}\right)$	$\Theta\left(\sqrt{M}\right)$

THEOREM 10 [CAPACITY FOR $M \overset{\lim}{<} KN$, $\mathcal{M}_\downarrow \neq \emptyset$]:

$$C = \begin{cases} \Theta\left(\sqrt{M}\right), & \textit{if } \tau < 1, \\ \Theta\left(\frac{\sqrt{M}}{\log M}\right), & \textit{if } \tau = 1, \\ \Theta\left(M^{\frac{3}{2}-\tau}\right), & \textit{if } 1 < \tau < \frac{3}{2}, \\ \Theta\left((\log r)^{\frac{3}{2}}\right), & \textit{if } \tau = \frac{3}{2}, \\ \Theta(1), & \textit{if } \tau > \frac{3}{2}. \end{cases}$$

THEOREM 11 [CAPACITY FOR $M \sim KN$, $\mathcal{M}_\downarrow \neq \emptyset$]:

$$C = \begin{cases} \Theta\left(\sqrt{M}\right), & \textit{if } \tau \leq 1, \\ \Theta\left(\frac{\sqrt{M}}{(KN-M)^{\tau-1}}\right), & \textit{if } 1 < \tau < \frac{3}{2}, \\ \Theta\left(\sqrt{\frac{M}{KN-M}}(\log r)^{\frac{3}{2}}\right), & \textit{if } \tau = \frac{3}{2}, \\ \Theta\left(\frac{\sqrt{M}}{(KN-M)^{\frac{3(\tau-1)}{2\tau}}}\right), & \textit{if } \tau > \frac{3}{2}. \end{cases}$$

C. Discussion on Asymptotic Laws

Table I summarizes the above results. C, the main result, is the minimum required link rate so as to be able to sustain a request rate $\lambda = 1$ from each node. Clearly, for the cases that C scales to infinity, as it is not possible to increase the link capacity to an arbitrary value, one should interpret the reported C as the inverse of the maximum sustainable request rate λ, e.g. the result of $C = \Theta\left((\log M)^{\frac{3}{2}}\right)$ with $\lambda = 1$ is equivalent to $C = 1$ with $\lambda = \Theta\left((\log M)^{-\frac{3}{2}}\right)$.

Regarding τ, note that there are two phase transition values, 1 and $3/2$, which lead to different scaling laws. The higher τ is, the higher the differentiation in the popularity of files, and thus, the better caching works, resulting in high performance (i.e. low link rate C). For example, for $\tau > \frac{3}{2}$ and $M \leq \delta N$, with $\delta < K$, it is $C = \Theta(1)$, meaning that the wireless network is indeed sustainable. Nevertheless, such a high τ distribution parameter is expected to appear only in cases that content is created from a particular service, e.g., mobile applications, as discussed in Section II.

In contrast, low values of τ (as in Internet traffic) flatten file popularity (i.e., $\tau = 0$ corresponds to the uniform distribution): then, replication cannot really help; we end up to $\Theta\left(\sqrt{M}\right)$, which is essentially the Gupta-Kumar link rates if we associate M, the number of files in our model, to N the number of communicating pairs in [2]. If M scales slower than N, then we get an improvement over [2], which however is attributed to the particular flow model and not to the effect of caching.

On the other hand, note that the case of M being a fraction of the total cache capacity KN, (e.g. $M \leq \delta KN$, with δ sufficiently small) is a reasonable regime if we make the following interpretation: each node brings to the network its own files and has enough spare capacity to cache files from other nodes. In such a case, M is the same order with N and, thus, the flow model is a fair comparison to [2]. The value of the multiplicative constant , however, dictates the behavior of the set $\mathcal{M}_{\downarrow}$ (from Theorem 7), and, consequently, the scaling law of C: $\Theta\left(\sqrt{M}\right)$ for $\tau < 1$, $\Theta\left(M^{\frac{3}{2}-\tau}\right)$ for $1 < \tau < \frac{3}{2}$, or $\Theta(1)$ for $\tau > \frac{3}{2}$, with the improvement here being attributed to the effect of finite caching. Finally, when the ratio of M over KN approximates 1, then there is little spare capacity for replication, therefore C in the last two columns of the tables essentially matches the Gupta-Kumar rate of $\Theta\left(\sqrt{M}\right)$.

VII. Conclusions & Future Work

In this paper, we investigated on the scaling properties in wireless networks with caching. In particular, we considered a square lattice topology, and derived the asymptotic laws of how link capacity scales up when request rate is constant (or inversely, how request rate should scale down when link capacity is constant) with the number of files M and number of nodes N. Our study shows that the file popularity parameter τ is the key factor on the impact of caching in the wireless network. An extended upcoming version of this work will include a new dimension of scaling with node capacity K; that is, in an evolving network, the amount of memory per node is envisaged to increase along with size of the network and the number of files in it. In such a setup, it is expected that caching can be beneficial even for low values of τ provided that cache capacity K scales sufficiently fast with the files M.

Acknowledgments

This work has been co-financed by the European Union (European Social Fund ESF) and Greek national funds through the Operational Program 'Education and Lifelong Learning of the National Strategic Reference Framework (NSRF) Research Funding Program': 'THALES–Investing in research and innovation through basic and applied research programs'. It has also been supported by the European Commission through the FP7 project STAMINA 265496.

Appendix

Proof of Lemma 2: If we assume that in the limit $l > 1$, then we have two cases for r: $r \to \infty$, or $r = O(1)$.

In the first case, $r \to \infty$ and $\tau \leq \frac{3}{2}$ lead to $H_{\frac{2\tau}{3}}(r-1)$ diverging to infinity in (11). However, the rest of the terms in (11) are bounded (as $l \leq K$). Therefore, in the limit, (11) is a contradiction. Thus, it has to be either $\mathcal{M}_{\updownarrow} = \emptyset$, or $l = 1$. As $r \to \infty$ and $l \leq K+1$, it cannot be $\mathcal{M}_{\updownarrow} = \emptyset$. Therefore, if $r \to \infty$ it is $l = 1$ (i.e., $d_l \ncong 1$ as assumed).

If, on the other hand, $r = O(1)$, then (15) is a contradiction for $l > 1$ in the limit. Thus, $l \to 1$. ∎

Proof of Lemma 3: Follows from the summation definition of (6): observe that $d_m > \frac{1}{N}$ and $\sum_{m \in \mathcal{M}_{\updownarrow}} p_m \leq 1$. ∎

Proof of Lemma 4: For $\tau < 1$, $C_{\downarrow} = \Theta\left(\sqrt{N}\right)$ follows from (7) and the fact that $H_\tau(M) - H_\tau(r) = \Theta\left(H_\tau(M)\right)$ from (4). The latter comes from $H_\tau(M)$ diverging and $r \overset{\lim}{<} M$.

For $\tau > 1$, it is $C_{\downarrow} \overset{(4)}{=} \Theta\left(\sqrt{N}\left[r^{1-\tau} - M^{1-\tau}\right]\right)$. If $r \overset{\lim}{<} M$, too, then $M^{1-\tau} \overset{\lim}{<} r^{1-\tau}$, hence $C_{\downarrow} = \Theta\left(\frac{\sqrt{N}}{r^{\tau-1}}\right)$. ∎

Proof of Theorem 6: **Case** $\tau \leq \frac{3}{2}$: From Lemma 2, $l \to 1$. **Case** $\tau > \frac{3}{2}$: Examining what happens in (9-10) in the limit, we observe that $r \to \infty$, hence $H_{\frac{2\tau}{3}}(r-1) \to H_{\frac{2\tau}{3}}$. Assuming a limit $l \to \hat{l}_{\{\tau > \frac{3}{2}, \mathcal{M} \approx \emptyset\}}$, (9-10) lead to (16); the latter can be shown (as in Theorem 1) to have a unique solution. ∎

Proof of Theorem 7: By the definition of $\mathcal{M}_{\downarrow}$ almost empty, it is $M - r + 1 = o(M)$, and given the constraint of (1), it is $M = O(N)$, thus $M - r + 1 = o(N) = o((K - l + 1)N)$, as $K - l + 1 \geq 1$ in all cases from Theorem 6.

From the last element of $\mathcal{M}_{\updownarrow}$, we have that $d_{r-1} > \frac{1}{N}$. Substituting d_{r-1} in the latter from (2b), and taking the limit

$$(K - l + 1)N > (r-1)^{\frac{2\tau}{3}} \left[H_{\frac{2\tau}{3}}(r-1) - H_{\frac{2\tau}{3}}(l-1)\right], \quad (24)$$

where we used $M - r + 1 = o((K - l + 1)N)$ to eliminate the respective term from the LHS. Next, we use (4) to approximate the Riemann terms and substitute l from Theorem 6:

Case $0 < \tau < \frac{3}{2}$: $l \to 1$, thus (24) becomes
$KN \overset{\lim}{\geq} (r-1)^{\frac{2\tau}{3}} \frac{(r-1)^{1-\frac{2\tau}{3}} - 1}{1 - \frac{2\tau}{3}} = \frac{r - 1 - (r-1)^{\frac{2\tau}{3}}}{1 - \frac{2\tau}{3}}$.
As $2\tau/3 < 1$, it is $(r-1)^{\frac{2\tau}{3}} = o(r-1)$. Hence, the above is equivalent in the limit to $(r-1) \overset{\lim}{\leq} K\left(1 - \frac{2\tau}{3}\right)N$, or, as $r = \Theta(M)$, $M \overset{\lim}{\leq} K\left(1 - \frac{2\tau}{3}\right)N$.

Case $\tau = \frac{3}{2}$: $l \to 1$, thus $KN \overset{\lim}{\geq} (r-1)\left[\ln(r-1) - \ln l\right]$.
Using that $\ln l = o(\ln(r-1))$, we get $(r-1)\ln(r-1) \overset{\lim}{\leq} KN$; as $r = \Theta(M)$, the condition becomes $M \ln M \overset{\lim}{\leq} KN$.

Case $\tau > \frac{3}{2}$:
$(K - \hat{l} + 1)N \overset{\lim}{\geq} (r-1)^{\frac{2\tau}{3}} \frac{\hat{l}^{1-\frac{2\tau}{3}} - (r-1)^{1-\frac{2\tau}{3}}}{\frac{2\tau}{3} - 1} = \frac{\hat{l}^{1-\frac{2\tau}{3}}(r-1)^{\frac{2\tau}{3}} - (r-1)}{\frac{2\tau}{3} - 1}$.

Now, it is $2\tau/3 > 1$, hence $(r-1) = o\left((r-1)^{\frac{2\tau}{3}}\right)$; then, the above becomes $(r-1) \overset{\lim}{\leq} \left[\frac{(K - \hat{l} + 1)\left(\frac{2\tau}{3} - 1\right)}{\hat{l}^{1-\frac{2\tau}{3}}} N\right]^{\frac{3}{2\tau}}$. Substituting $r - 1$ with M, the condition follows.

Last, observe that in the above derivations, if we started with $d_M > 1/N$, then we would find the conditions for $\mathcal{M}_{\downarrow}$ being strictly empty, i.e. $r = M + 1$. As easily seen, this is true if the conditions are satisfied with strict inequality. ∎

Proof of Theorem 8: To find C, we compute $C_{\updownarrow}$ and $C_{\downarrow}$ from their definitions (6-7), and show that in all cases $C_{\downarrow} = O\left(C_{\updownarrow}\right)$. Thus, $C = \Theta\left(C_{\updownarrow}\right)$. In the computation of $C_{\downarrow}$,

$r \sim M$ helps in deriving that

$$H_\tau(M) - H_\tau(r-1) = \sum_{j=r}^{M} j^{-\tau} = \Theta(M^{-\tau}(M-r))\,.$$

Theorem 6 and $M - r = o(M)$ result in $K_{\circ} = \Theta(1)$ for all τ.
Case $\tau < 1$: Regarding $C_{\circ}$, $H_{\frac{2\tau}{3}}(r+1)$ and $H_\tau(M)$ diverge, while $H_{\frac{2\tau}{3}}(l-1)$ is bounded (as $l \leq K+1$). Thus,

$$C_{\circ} = \Theta\left(\frac{\left[M^{1-\frac{2\tau}{3}}-1\right]^{\frac{3}{2}}}{M^{1-\tau}-1}\right) = \Theta\Big(\sqrt{M}\Big)\,.$$

If the condition of Theorem 7 is a strict inequality, then $C_{\bullet} = 0$. If it is an equality, it is $M = \Theta(N)$. Then,

$$C_{\bullet} = \sqrt{N}\frac{\sum_{j=r}^{M} j^{-\tau}}{H_\tau(M)} \overset{(4)}{=} \Theta\Big(\sqrt{N}\frac{M^{-\tau}(M-r)}{M^{1-\tau}}\Big) \overset{M=\Theta(N)}{=} o\Big(\sqrt{M}\Big).$$

Case $\tau = 1$: In $C_{\circ}$, $H_{\frac{2\tau}{3}}(M)$ and $H_\tau(M)$ diverge, while $H_{\frac{2\tau}{3}}(l-1)$ is bounded. Thus, $C_{\circ} = \Theta\left(\frac{\left[M^{\frac{1}{3}}-1\right]^{\frac{3}{2}}}{\log M}\right) = \Theta\Big(\frac{\sqrt{M}}{\log M}\Big)$.

As before, if the condition of Theorem 7 is a strict inequality, then $C_{\bullet} = 0$. If it is an equality, it is $M = \Theta(N)$. Then,

$$C_{\bullet} = \sqrt{N}\frac{\sum_{j=r}^{M} j^{-1}}{H_\tau(M)} = \Theta\Big(\sqrt{N}\frac{M^{-1}(M-r)}{\log M}\Big) \overset{M=\Theta(N)}{=} o\Big(\frac{\sqrt{M}}{\log M}\Big)\,.$$

Case $1 < \tau < \frac{3}{2}$: Regarding $C_{\circ}$, only $H_{\frac{2\tau}{3}}(M)$ diverges, while the rest of the terms converge. Then, the order of $C_{\circ}$ is determined from

$$H^{\frac{3}{2}}_{\frac{2\tau}{3}}(M) \sim \left[\frac{M^{1-\frac{2\tau}{3}}-1}{1-\frac{2\tau}{3}}\right]^{\frac{3}{2}} = \Theta\Big(M^{\frac{3}{2}-\tau}\Big)\,.$$

If the condition of Theorem 7 is a strict inequality, then $C_{\bullet} = 0$. Otherwise, it is an equality, thus $M = \Theta(N)$, and

$$C_{\bullet} = \sqrt{N}\frac{\sum_{j=r}^{M} j^{-\tau}}{H_\tau(M)} \overset{(4)}{=} \Theta\Big(\sqrt{N}M^{-\tau}(M-r)\Big) \overset{M=\Theta(N)}{=} o\Big(M^{\frac{3}{2}-\tau}\Big)\,.$$

Case $\tau = \frac{3}{2}$: $C_{\circ} = \Theta\Big((\log M)^{\frac{3}{2}}\Big)$ due to the numerator, all other terms converge. If the condition of Theorem 7 is a strict inequality, $C_{\bullet} = 0$. Otherwise, it is an equality with $M = \Theta(N)$, and given that $1 < \tau < \frac{3}{2}$, $C_{\bullet} = \Theta\Big(\sqrt{N}\frac{M-r}{M^{\frac{3}{2}}}\Big) = \Theta\big(\frac{M-r}{M}\big) = o(1)$. In total, $C = \Theta\Big((\log M)^{\frac{3}{2}}\Big)$.

Case $\tau > \frac{3}{2}$: All terms converge in (6), thus $C_{\circ} = O(1)$. If the condition of Theorem 7 is a strict inequality, $C_{\bullet} = 0$. Otherwise, it is an equality with $M = \Theta\left(N^{\frac{3}{2\tau}}\right)$. Then, $C_{\bullet} = \Theta\Big(\sqrt{N}\frac{M-r}{M^\tau}\Big) = \Theta\Big(\frac{M-r}{M^{\frac{2\tau}{3}}}\Big) = o(1)$. In total, $C = O(1)$. ■

Proof of Theorem 9: In the second case of $KN - M = O(1)$, observe that $KN - M$ is the number of places that remain after storing the M files once in the network. It is easy to see that $r \leq KN - M$, thus $r = O(1)$. As both r and l are bounded, (15) cannot be true, therefore $\hat{l} = 1$. Hence, r is estimated from (12), (13); substituting $l \to 1$, (23) follows.

For the first part where $KN - M = \omega(1)$, we first note that $r = O(N)$, as $r \leq M$, and $M = O(N)$, due to (1).
Case $\tau < \frac{3}{2}$: From Lemma 2, $l \to \hat{l} = 1$. Using this along with (4) in (14), we can estimate r:

$$KN - M + r - 1 \cong (r-1)^{\frac{2\tau}{3}}\frac{r^{1-\frac{2\tau}{3}}-1}{1-\frac{2\tau}{3}}$$

Observe that assuming $r = O(1)$, the above results becomes a contradiction, as $NK - M = \omega(1)$, whereas all the other terms are $O(1)$. Therefore, it is $r = \omega(1)$, and (18) follows.
Case $\tau = \frac{3}{2}$: From Lemma 2, $l \to \hat{l} = 1$. Working as before, (14) in view of (4) gives that $NK - M + r - 1 \cong (r-1)\ln r$. Clearly, $r = \omega(1)$, thus, $r \ln r \sim KN - M$.
Case $\tau > \frac{3}{2}$: First, let us assume that $\hat{l} > 1$, and use (11), which using (4) is approximated by

$$K - l + 1 - \frac{M - r + 1}{N} \cong l^{\frac{2\tau}{3}}\frac{l^{1-\frac{2\tau}{3}} - r^{1-\frac{2\tau}{3}}}{\frac{2\tau}{3}-1} \cong \frac{l - l^{\frac{2\tau}{3}}r^{1-\frac{2\tau}{3}}}{\frac{2\tau}{3}-1} \Rightarrow$$

$$K - l + 1 - \frac{M}{N} + \frac{l}{N^{1-\frac{3}{2\tau}}} \cong l\frac{1-\frac{1}{N^{1-\frac{3}{2\tau}}}}{\frac{2\tau}{3}-1},$$

where in the last step we used (15) to substitute $r \cong lN^{\frac{3}{2\tau}}$. For $N \to \infty$, it is $N^{1-\frac{3}{2\tau}} \to \infty$, and the above becomes

$$\hat{l} \cong \frac{2\tau - 3}{2\tau}\left(K + 1 - \lim\frac{M}{N}\right).$$

Thus, the assumption of $\hat{l} > 1$ is correct if $K, \lim M/N$ and τ are such that the second factor of RHS is approximately exceeds 1, that is $M \overset{\lim}{\leq} \left(K - \frac{3}{2\tau-3}\right)N$. Then, from (15), r is:

$$r \sim \frac{2\tau - 3}{2\tau}\left[(K+1)N^{\frac{3}{2\tau}} - \frac{M}{N^{1-\frac{3}{2\tau}}}\right].$$

Otherwise, $\hat{l} = 1$, and r is computed from (14) using (4)

$$NK - M + r - 1 \cong (r-1)^{\frac{2\tau}{3}}\frac{1-r^{1-\frac{2\tau}{3}}}{\frac{2\tau}{3}-1}.$$

As $N \to \infty$, it follows that $r \sim \left[\frac{2\tau-3}{3}(KN - M)\right]^{\frac{3}{2\tau}}$. ■

Proof of Theorem 10: First note that from Theorem 9, for all τ, it is $K_{\circ} = \Theta\left(\frac{KN - M + \hat{r} - 1}{N}\right) = \Theta(1)$ (using $M \overset{\lim}{<} KN$).

In the cases of $\tau < \frac{3}{2}$, $\mathcal{M}_{\bullet} \neq \emptyset$ entails $M \overset{\lim}{>} K\left(1 - \frac{2\tau}{3}\right)N$ (Theorem 7). It is also $M \overset{\lim}{<} KN$, thus, $M = \Theta(N)$.

Furthermore, from Theorem 9 and $M \overset{\lim}{<} KN$, it is

$$r \sim \frac{3 - 2\tau}{2\tau}(KN - M) + 1 \overset{M \overset{\lim}{\leq} KN}{=} \Theta(N), \text{ and, moreover,}$$

$$r \overset{\lim}{<} \frac{3-2\tau}{2\tau}\,\frac{2\tau}{3}KN = \left(1 - \frac{2\tau}{3}\right)KN \overset{\lim}{<} M. \quad (25)$$

Then, we compute the link rate as follows:
Case $\tau < 1$: Using Lemma 4, it is $C_{\bullet} = \Theta\left(\sqrt{N}\right)$. Invoking Lemma 3, too, we get that $C = \Theta\Big(\sqrt{N}\Big) = \Theta\Big(\sqrt{M}\Big)$.
Case $\tau = 1$: $C_{\bullet} = \sqrt{N}\frac{H_1(M) - H_1(r)}{H_1(M)} \overset{(4)}{\sim} \sqrt{N}\frac{\ln\frac{M}{r}}{\ln M} = \Theta\Big(\frac{\sqrt{M}}{\log M}\Big)$,

using $r = \Theta(N) = \Theta(M)$. Similarly, as $l \to 1$, $C_{\circ} = \Theta\left(\frac{H^{\frac{3}{2}}_{\frac{2}{3}}(r-1)}{\sqrt{K_{\circ}}\;H_1(M)}\right) = \Theta\Big(\frac{\sqrt{N}}{\log N}\Big)$. In total, $C = \Theta\Big(\frac{\sqrt{M}}{\log M}\Big)$.
Case $1 < \tau < \frac{3}{2}$: Using $r = \Theta(N) = \Theta(M)$,

$$C_{\downarrow} = \sqrt{N}\ \frac{H_\tau(M) - H_\tau(r)}{H_\tau(M)} \overset{(4)}{\sim} \frac{\sqrt{M}}{r^{\tau-1}}\left[1-\left(\frac{r}{M}\right)^{\tau-1}\right],$$

which is $C_{\downarrow} = O\left(M^{\frac{3}{2}-\tau}\right)$ from (25). Last, $l \to 1$ implies that $C_{\updownarrow} \sim \frac{H^{\frac{3}{2}}_{\frac{2\tau}{3}}(r-1)}{\sqrt{K_{\updownarrow}}\ H_\tau(M)} = \Theta\left(M^{\frac{3}{2}-\tau}\right)$. In total, $C = \Theta\left(M^{\frac{3}{2}-\tau}\right)$.

Case $\tau = \frac{3}{2}$: Now, it has to be $M \ln M \overset{\lim}{>} KN$, which also implies that $M \log M = \Omega(N)$. From Theorem 9, we have that $r \ln r \sim KN - M$. This means that $r \log r = \Theta(N)$ in view of $M \overset{\lim}{<} KN$, and thus $r = o(N)$.

Moreover, comparing $M \ln M$ and $r \ln r$ in the above formulas, it has to be $r \overset{\lim}{<} M$. The latter implies that there exists a $0 < k < 1$ such that eventually $\frac{r}{M} \le k$. Using then (7),

$$\begin{aligned} C_{\downarrow} &\overset{(4)}{=} \Theta\left(N^{\frac{1}{2}}\left[\frac{1}{r^{\frac{1}{2}}} - \frac{1}{M^{\frac{1}{2}}}\right]\right) \overset{\left[\frac{r}{M}\right]^{\frac{1}{2}} \le \sqrt{k}}{=} \Theta\left(\sqrt{\frac{N}{r}}\right) \\ &= \Theta\left(\sqrt{\frac{N}{KN-M}\log r}\right) \overset{NK-M=\Theta(N)}{=} \Theta\left(\sqrt{\log r}\right). \end{aligned}$$

Moreover, as $l \to 1$, $C_{\updownarrow} = \Theta\left(\frac{H_1^{\frac{3}{2}}(r)}{\sqrt{K_{\updownarrow}}\ H_{\frac{3}{2}}(M)}\right) = \Theta\left((\log r)^{\frac{3}{2}}\right)$.

Thus, in total $C = \Theta\left((\log r)^{\frac{3}{2}}\right)$.

Case $\tau > \frac{3}{2}$: it is $r = \Theta\left(N^{\frac{3}{2\tau}}\right)$ due to $M \overset{\lim}{<} KN$. Moreover, as for $\mathcal{M}_{\downarrow} \neq \emptyset$, it has to be $M = \Omega\left(N^{\frac{3}{2\tau}}\right)$ for $\mathcal{M}_{\downarrow}$. Then,

$$\begin{aligned} C_{\downarrow} &= \sqrt{N}\frac{H_\tau(M) - H_\tau(r-1)}{H_\tau(M)} = O\left(N^{\frac{1}{2}} r^{1-\tau}\right) \\ &= O\left(N^{\frac{1}{2}+\frac{3}{2\tau}(1-\tau)}\right) = O\left(N^{\frac{3}{2\tau}-1}\right) = O(1). \end{aligned}$$

Last, $C_{\updownarrow} = \Theta(1)$ (all terms converge). Thus, $C = \Theta(1)$. ∎

Proof of Theorem 11: In all the cases, we know that $r \le KN - M + 1$, as $KN - M$ is the number of spaces left for duplicate copies after all M files are stored once. Hence, $r = O(KN - M) = o(N) = o(M)$. Moreover, as before, in all cases, $K_{\updownarrow} = \Theta\left(\frac{KN-M+r-1}{N}\right) = \Theta\left(K - \frac{M}{N}\right)$.

Case $\tau \le 1$: From Lemma 4, $r = \omega(M)$ implies that $C_{\downarrow} = \Theta\left(\sqrt{N}\right)$. Hence, invoking Lemma 3, $C \overset{M=\Theta(N)}{=} \Theta\left(\sqrt{M}\right)$.

For the rest of the cases with $\tau > 1$, it is $r = o(M)$, therefore, from Lemma 4, we get that $C_{\downarrow} = \Theta\left(\frac{\sqrt{N}}{r^{\tau-1}}\right)$.

Case $1 < \tau < \frac{3}{2}$: Using $r = \Theta(KN - M)$ from Theorem 9, $C_{\downarrow} = \Theta\left(\frac{\sqrt{N}}{(KN-M)^{\tau-1}}\right)$. On the other hand, $l \to 1$, and thus $C_{\updownarrow} = \frac{H^{\frac{3}{2}}_{\frac{2\tau}{3}}(r-1)}{\sqrt{K_{\updownarrow}} H_\tau(M)} = \Theta\left(\sqrt{\frac{N}{KN-M}} r^{\frac{3}{2}-\tau}\right) = O\left(\frac{\sqrt{N}}{(KN-M)^{\tau-1}}\right)$.

In total, $C \overset{M=\Theta(N)}{=} \Theta\left(\frac{\sqrt{M}}{(KN-M)^{\tau-1}}\right)$.

Case $\tau = \frac{3}{2}$: From the above, $C_{\downarrow} = \Theta\left(\sqrt{\frac{N}{r}}\right)$. Moreover,

$$C_{\updownarrow} = \frac{H_1^{\frac{3}{2}}(r-1)}{\sqrt{K_{\updownarrow}} H_{\frac{3}{2}}(M)} = \Theta\left(\sqrt{\frac{N}{KN-M}} \log^{\frac{3}{2}} r\right).$$

However, $\frac{1}{r} = \frac{\log r}{r \log r} = \Theta\left(\frac{\log r}{KN-M}\right) = o\left(\frac{\log^3 r}{KN-M}\right)$, thus $C_{\downarrow} = o(C_{\updownarrow})$. In total, $C \overset{M=\Theta(N)}{=} \Theta\left(\sqrt{\frac{M}{KN-M}} \log^{\frac{3}{2}} r\right)$.

Case $\tau > \frac{3}{2}$: From Theorem 9, $r = \Theta\left((KN-M)^{\frac{3}{2\tau}}\right) = o(KN - M) = o(M)$. Thus, $C_{\downarrow} = \Theta\left(\frac{\sqrt{N}}{(KN-M)^{\frac{3}{2}\ \frac{\tau-1}{\tau}}}\right)$. Moreover, $C_{\updownarrow} = \Theta\left(K_{\updownarrow}^{-\frac{1}{2}}\right) = \Theta\left(\sqrt{\frac{N}{KN-M}}\right)$ (the H-terms converge). As $\frac{3(\tau-1)}{2\tau} > \frac{1}{2}$, it is $C \overset{M=\Theta(N)}{=} \Theta\left(\sqrt{\frac{M}{KN-M}}\right)$. ∎

REFERENCES

[1] V. Jacobson, D. K. Smetters, J. D. Thornton, M. F. Plass, N. H. Briggs, and R. L. Braynard, "Networking named content," in *Proc. of the 5th international conference on Emerging networking experiments and technologies (CoNEXT '09)*, Rome, Italy, Dec. 2009, pp. 1–12.

[2] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 46, pp. 388–404, Mar. 2000.

[3] A. Özgür, O. Lévêque, and D. Tse, "Hierarchical cooperation achieves optimal capacity scaling in ad hoc networks," *IEEE Trans. Inf. Theory*, vol. 53, pp. 3549–3572, Oct. 2007.

[4] M. Franceschetti, M. D. Migliore, and P. Minero, "The capacity of wireless networks: information-theoretic and physical limits," *IEEE Trans. Inf. Theory*, vol. 55, pp. 3413–3424, Aug. 2009.

[5] S. Toumpis, "Asymptotic capacity bounds for wireless networks with non-uniform traffic patterns," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 2231–2242, Jun. 2008.

[6] A. Zemlianov and G. de Veciana, "Capacity of ad hoc wireless networks with infrastructure support," *IEEE J. Sel. Areas Commun.*, vol. 23, pp. 657–667, Mar. 2005.

[7] S. Jin and L. Wang, "Content and service replication strategies in multi-hop wireless mesh networks," in *MSWiM '05: Proceedings of the 8th ACM international symposium on Modeling, analysis and simulation of wireless and mobile systems*, Montréal, Quebec, Canada, Oct. 2005, pp. 79–86.

[8] V. Sourlas, G. S. Paschos, P. Flegkas, and L. Tassiulas, "Mobility support through caching in content-based publish-subscribe networks," in *Proc. of 5th International workshop on content delivery networks (CDN 2010)*, Melbourne, Australia, May 2010, pp. 715–720.

[9] J. Zhao, P. Zhang, G. Cao, and C. R. Das, "Cooperative caching in wireless p2p networks: Design, implementation, and evaluation," *IEEE Trans. Parallel Distrib. Syst.*, vol. 21, pp. 229–241, Feb. 2010.

[10] U. Niesen, D. Shah, and G. Wornell, "Caching in wireless networks," in *IEEE International Symposium on Information Theory*, Seoul, Korea, Jun. 2009, pp. 2111–2115.

[11] J. Silvester and L. Kleinrock, "On the capacity of multihop slotted aloha networks with regular structure," *IEEE Trans. Commun.*, vol. 31, pp. 974–982, Aug. 1983.

[12] M. E. J. Newman, "Power laws, pareto distributions and Zipf's law," *Contemporary Physics*, vol. 46, pp. 323–351, Sep./Oct. 2005.

[13] L. A. Adamic and B. A. Huberman, "Zipf's law and the Internet," *Glottometrics*, vol. 3, pp. 143–150, 2002.

[14] J. Chu, K. Labonte, and B. N. Levine, "Availability and popularity measurements of peer-to-peer file systems," in *Proceedings of SPIE*, Boston, MA, USA, Jul. 2002.

[15] T. Yamakami, "A Zipf-like distribution of popularity and hits in the mobile web pages with short life time," in *Proc. of Parallel and Distributed Computing, Applications and Technologies, PDCAT '06*, Taipei, ROC, Dec. 2006, pp. 240–243.

[16] L. Breslau, P. Cue, P. Cao, L. Fan, G. Phillips, and S. Shenker, "Web caching and Zipf-like distributions: Evidence and implications," in *Proc. of INFOCOM*, New York, NY, USA, Mar. 1999, pp. 126–134.

[17] C. R. Cunha, A. Bestavros, and M. E. Crovella, "Characteristics of WWW Client-based Traces," in *View on NCSTRL*, Boston University, MA, USA, Jul. 1995.

[18] M. Franceschetti and R. Meester, *Random Networks for Communication*. New York, NY, USA: Cambridge University Press, Series: Cambridge Series in Statistical and Probabilistic Mathematics (No. 24), 2007.

[19] S. Gitzenis, G. S. Paschos, and L. Tassiulas, "*Asymptotic Laws for Joint Content Replication and Delivery in Wireless Networks*," in *arxiv.org* on-line repository, Jan. 2012.