HAL
open science

# The Effect of Caching in Sustainability of Large Wireless Networks

Georgios Paschos, Savvas Gitzenis, Leandros Tassiulas

## To cite this version:

Georgios Paschos, Savvas Gitzenis, Leandros Tassiulas. The Effect of Caching in Sustainability of Large Wireless Networks. WiOpt'12: Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks, May 2012, Paderborn, Germany. pp.355-360. hal-00766458

**HAL Id: hal-00766458**

**https://inria.hal.science/hal-00766458**

Submitted on 18 Dec 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# The Effect of Caching in Sustainability of Large Wireless Networks

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

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

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

***Abstract*—We study the scalability of multihop wireless communications, a major concern in networking, for the case that users access content cached across the nodes. In contrast to the standard paradigm of randomly selected communicating pairs, content replication is efficient for certain regimes of content volume and popularity, cache and network size. Assuming the Zipf popularity law, and investigating on the relative ways that the number of files, the cache size and the network nodes can all jointly scale to infinity, we derive asymptotic laws about the required link capacity, which range from $O\left(\sqrt{N}\right)$ down to $O(1)$, and identify the associated regimes of network operation.**

## I. Introduction

The proliferation of video applications and the advent of new paradigms like multiview and 3D video, as well as other demanding applications push the operation of networks to their physical limits. To overcome these challenges, the networking community devises new technologies and architectures like the Peer-to-Peer (P2P) communication and the Content Centric Networking in an effort to improve the scalability and the efficiency for the future networks. In this landscape, wireless networks are considered to hold an important role, supporting mobility of users, extending network connectivity and promoting ubiquitous computing. By 2015, traffic from wireless devices will exceed the traffic from the wired domain [1].

Despite their worldwide deployment, wireless networks are mostly confined tof wireless networks is limited to specific domains like sensor networks where the supported communication rates are low, or engineered fixed point-to-point links with directional antennas. In the seminal work of Gupta-Kumar [2], the traffic-carrying capacity of a planar wireless network was shown to be $O\left(\frac{1}{\sqrt{N}}\right)$, where $N$ is the number of nodes. This implies that large multihop networks cannot sustain throughput among random pairs due to the increasing hop number between the source and the destination.

There is anecdotal evidence that the average P2P file travels on the same optical link multiple times. Taking into account that P2P video currently accounts for almost 50% of the overall network traffic [1], the impact in network efficiency is significant. In this context, *network caching* can play a key role in mitigating these inefficiencies and avoiding excess traffic by storing the data close to the customer.

In the future, one can envisage myriads of wireless devices with significant caching capabilities forming novel multihop wireless networks, and delivering high quality services as the already described. In particular, the size of these caches is expected to increase geometrically[1] over time. Thus, the emerging question is *whether a measurable improvement in the operation of wireless networks can come from the growth in cache sizes*, and, in particular, in the asymptotic law regimes.

In this work, we depart from the random-pairs paradigm of [2] and important follow-ups (e.g., [4]–[8]), by adopting a content-based network, where nodes request content replicated in the caches of other nodes. This paradigm gives rise to the joint problem of *replication* and *routing*. Assuming the symmetric topology of the square grid—a well-accepted model for various planar wireless networks, we set up a simple replication problem whose optimal solution results in the same order to the complex combinatorial joint problem. Then, we use this solution to identify the asymptotic laws of the required link capacity that can sustain the associated network traffic.

In contrast to our prior works [9], [10], the cache size is assumed to increase to infinity. In such a study, the statistics of the applications are quite important. We, therefore, assume a popularity about the requested messages described by the Zipf law, a model well-known regarding Internet traffic.

Due to space constraints, the proofs will be appearing in an upcoming extended version. Table I provides the definitions on the asymptotic notation that we use throughout this work.

## II. Basic Definitions and the Density Problem

Assume a square lattice with $N$ nodes, with $N$ being the square of an integer; the set of nodes is indexed by $n \in \mathcal{N} \triangleq \{1, 2, \ldots, N\}$. Each node is connected to its four neighbors that lie next to it on the same row or column with undirected links. A scaling network emerges by keeping the node density fixed and increasing the network size $N$, as in [2]. To avoid boundary effects, we consider a toroidal structure as in [11].

Nodes (or users located therein) generate requests to access files/data, indexed by $m \in \mathcal{M} \triangleq \{1, 2, \ldots, M\}$. Each node $n$ is equipped with a cache/buffer, whose contents are denoted by the set $\mathcal{B}_n$, a subset of $\mathcal{M}$. If a request at node $n$ regards a file $m$ that lies in $\mathcal{B}_n$, then it is served locally. Due to the limited buffer capacity, $m$ will often be not available locally,

[1]In the last couple of decades, after a period of doubling of the areal density of hard disks per year, the growth rate has dropped to doubling every three years. Similarly, DRAM capacity quadruples every three years [3].

978-1-61284-824-2/2012 - Copyright is with IFIP

thus, node $n$ will have to retrieve $m$ over the network from some other node(s) $w$ that keep $m$ in its cache. Thus for each $n, m$ pair, a route (or set of) $\mathcal{R}_{n,m}$ should be specified for the path(s) followed from $n$ to one or more $w$ to retrieve $m$.

Let, moreover, $K$ be the storage capacity of nodes' cache measured in the number of files it can store. This means that all $M$ files are of the same (unit) size, placing a constraint on the cardinality of cache contents $|\mathcal{B}_n| \leq K$. The generalization to variable sized files can be still captured in this framework by splitting each large file into multiple unit segments, and then treating its segments as separate, independent files.

For the problem to be non-trivial and feasible, it should be

$$K < M \leq KN. \tag{1}$$

The first inequality implies that each node has to select the files of its cache, and the second is equivalent to the network having sufficient capacity to store all files at least once.

Last, let each node $n \in \mathcal{N}$ generate i.i.d. requests[2] for data at rate $\lambda_n$. In this work, we focus on symmetric request rates, that is, $\lambda_n = \lambda = 1$. Each request regards a particular file $m \in \mathcal{M}$, depending on the file $m$'s *popularity* $p_m$. In essence, $[p_m]$ is a probability distribution, i.e., sets the probability of a request for a given file. Clearly, replication should be governed by the popularity: storing the popular files densely will maximize the gain of caching on the network traffic.

### A. General Replication-Delivery Problem

Assuming a $[\mathcal{B}_n]$ replication and delivery routes $[\mathcal{R}_{n,m}]$, it is easy to find the traffic load $C_\ell$ at each link $\ell$ of the network. The associated replication-delivery problem [9], [10] regards minimizing over the worst (or, in a relaxed form, average) traffic $C_\ell$ given the constraints of (i) node capacity and (ii) storing at least one copy of each file over the network. The resulting $C_\ell$, then, sets the minimum *capacity* of each link so that the network operates properly (i.e., be stable).

Obviously, the entanglement between buffer contents $[\mathcal{B}_n]$ (where each file is stored) and the delivery paths $[\mathcal{R}_{n,m}]$ calls for a joint optimization. This is clearly a combinatorial complexity problem, and thus not amenable to an easy-to-compute solution. Therefore, we should seek suitable simplifications and approximations to derive a suboptimal but efficient solution. For the needs of our study, this translates to an order-optimal solution, i.e., the proposed suboptimal solution lies within a constant to the optimal (but hard to compute) solution.

### B. Replication Density-based Problem

Given a $[\mathcal{B}_n]$ assignment, consider the frequency of occurrence of each file $m$ in the caches, or *replication density* $d_m$ as the fraction of nodes that store file $m$ in the network:

$$d_m = \frac{1}{N} \sum_{n \in \mathcal{N}} \mathbb{1}_{\{m \in \mathcal{B}_n\}}.$$

Based on this metric, we define a much simpler problem:

**PROBLEM 1:** *Minimize* $C \triangleq \sum_{m \in \mathcal{M}} \left[ \frac{1}{\sqrt{d_m}} - 1 \right] p_m$, *s.t.*

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

In the above, the optimization variables are the densities $d_m$, which express the fraction of caches containing file $m$. In the objective, $d_m^{-\frac{1}{2}} - 1$ approximates the average hop count from a random node to a cache containing $m$. Weighted by the probability $p_m$ of requests on $m$, the summation expresses the average link load per request.

This optimization is shown in [10] to be a relaxed version of the actual general problem; moreover, its optimal solution $C$ is of the same order to the solution of the original problem. In particular, [10] presents an order-optimal algorithm to assign the node cache contents $[\mathcal{B}_n]$ from the densities $d_m$ and uses shortest path routing for the delivery paths $[\mathcal{R}_{n,m}]$. Thus, the asymptotic laws of the original problem and of $C$ coincide, and, therefore, it suffices to focus on $C$'s scaling.

It should be noted that a similar optimization is formulated in [12], without, however, the $\frac{1}{N} \leq d_m \leq 1$ constraints. As seen next, these inequalities play a major role on the problem and its solution, and thus have a key effect in the asymptotics.

### C. Density Problem Solution

As explained in [9], [10], and easily seen from the functional form, the density problem admits a unique solution using the Karush-Kuhn-Tucker (KKT) conditions, and a computationally efficient algorithm to find it in polynomial time. Regarding the constraints on $d_m$ about its minimum and maximum value, either one of them can be an equality, or none. This induces a partition of $\mathcal{M}$ into three subsets, one containing files of unit replication density (i.e., stored at every node) $\mathcal{M}_{\uparrow} = \{m : d_m = 1\}$, one containing files stored in just one node $\mathcal{M}_{\downarrow} = \{m : d_m = \frac{1}{N}\}$, and the complementary $\mathcal{M}_{\updownarrow} = \mathcal{M} \setminus (\mathcal{M}_{\uparrow} \cup \mathcal{M}_{\downarrow})$ of files with $\frac{1}{N} < d_m < 1$.

When $p_m$ is in decreasing order, these sets become ordered, too, as illustrated in Fig. 1; then, we use variables $l$ and $r$ to identify the boundaries of these sets, as follows: $\mathcal{M}_{\uparrow} = \{1, 2, \ldots, l-1\}$, $\mathcal{M}_{\updownarrow} = \{l, l+1, \ldots, r-1\}$, and $\mathcal{M}_{\downarrow} = \{r, r+2, \ldots, M\}$, where $l$ and $r$ are integers with $1 \leq l \leq r \leq M+1$. Given these, the solution $d_m$ takes the form of

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

## III. Asymptotic Laws for Zipf Popularity

To derive concrete asymptotics about $C$, we consider Zipf, a well-known law regarding the traffic in the Internet.

### A. Zipf Law and Approximations

The Zipf distribution is defined as follows:

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

where $\tau$ is the power law parameter, adjusting the rate of

[2]No spatial or temporal dependence on the users' requests is assumed.

TABLE I
DEFINITION OF ASYMPTOTIC NOTATION ($f$ and $g$ are positive functions).

| | | |
|---|---|---|
| $f = o(g)$ | For any $k > 0$, there exists $\hat{x}$: | $x \geq \hat{x} \Rightarrow f(x) \leq kg(x)$ |
| $f = O(g)$ | There exists $k > 0$ and $\hat{x}$: | |
| $f \overset{\lim}{\leq} kg$ | There exists $\hat{x}$: | |
| $f \overset{\lim}{<} k'g$ | For any $0<k<k'$, there exists $\hat{x}$: | |
| $f = \omega(g)$ | For any $k > 0$, there exists $\hat{x}$: | $x \geq \hat{x} \Rightarrow f(x) \geq kg(x)$ |
| $f = \Omega(g)$ | There exists $k > 0$ and $\hat{x}$: | |
| $f \overset{\lim}{\geq} kg$ | There exists $\hat{x}$: | |
| $f \overset{\lim}{>} k'g$ | For any $k > k'$, there exists $\hat{x}$, | |
| $f = \Theta(g)$ | Iff $f = O(g)$ and $f = \Omega(g)$ | |

popularity decline with $m$; $H_\tau(n) \triangleq \sum_{j=1}^{n} j^{-\tau}$ is the truncated zeta function evaluated at $\tau$ (also called the $n^{\text{th}}$ *$\tau$-order generalized harmonic number*). The limit $H_\tau \triangleq \lim_{n\to\infty} H_\tau(n)$ is the Riemann zeta function, which converges when $\tau > 1$. To approximate $H_\tau(n)$, we bound the sum: for $n \geq m \geq 0$,

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

For any $m < n$ such that $n \sim m$, we will make use of the following approximation derived by counting the sum terms

$$H_\tau(n) - H_\tau(m) = \sum_{j=m}^{n} j^{-\tau} \sim n^{-\tau}(n-m). \tag{5}$$

Plugging the Zipf distribution and (2) into $C$, it follows that

$$C \triangleq \sum_{m\in\mathcal{M}} \left(d_m^{-\frac{1}{2}} - 1\right) p_m = C_{\text{⇂}} + C_{\text{↓}} - \sum_{j=l}^{M} p_m,$$

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

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

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

Next, we let all $K, N, M$ go to infinity and derive the scaling laws of $C$ (summarized in Table II). First, however, we have to derive approximations about the indices $l, r$, as they are not provided in closed form.

### B. Estimation of $l$ and $r$

*1) Estimation of $l$:* if $\mathcal{M}_{\text{⇂}} \cup \mathcal{M}_{\text{↓}}$ is not empty, $d_l < 1 \overset{(2b)}{\Leftrightarrow}$

$$K-l+1-\frac{M-r+1}{N} < l^{\frac{2\tau}{3}}\left[H_{\frac{2\tau}{3}}(r-1) - H_{\frac{2\tau}{3}}(l-1)\right]. \tag{8}$$

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

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

Thus, $l$ can be uniquely determined as the lowest integer that satisfies (8)-(9). Thus, $l$ can be estimated by treating (8) as an approximate equality (as $d_{l-1} = 1$ and $d_l < 1$):

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

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

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

Again, if set $\mathcal{M}_{\text{↓}}$ is not empty, i.e., $r \leq M$, then $d_r = N^{-1}$. Thus, if we attempted increasing index $r$ by one, (2b) would violate the constraint resulting in a density less than $N^{-1}$:

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

As before, (11) is an approximate equality, i.e.,

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

*3) Estimation of $\frac{l}{r}$:* For all $l, r$, it is $N > \frac{d_l}{d_{r-1}} = \left(\frac{r-1}{l}\right)^{\frac{2\tau}{3}}$. As before, whenever $l$ and $r$ are not equal to the extremes, i.e., $1 < l \leq r < M+1$, it holds $d_{l-1}/d_r = N$. Hence,

$$l \cong rN^{-\frac{3}{2\tau}}. \tag{14}$$

Up to this point, we have summarized the analysis of [9], [10]. Next, we let $N, M$ and $K$ jointly deviate to infinity.

### C. $l$ and $r$ on Almost Empty $\mathcal{M}_{\text{↓}}$

The first case of interest is when the solution results in an almost empty set $\mathcal{M}_{\text{↓}}$. Formally, $\mathcal{M}_{\text{↓}} \approx \emptyset$ iff $|\mathcal{M}_{\text{↓}}| = o(M)$, i.e., the number of last set's elements over the total files is negligible; thus, $\mathcal{M}_{\text{↓}} = \emptyset$ is a special case of $\mathcal{M}_{\text{↓}} \approx \emptyset$.

For $\mathcal{M}_{\text{↓}} \approx \emptyset$, $M$ should increase at a slow pace with respect to $N$ and $K$, so that the constraint $d_m \geq N^{-1}$ is satisfied for almost all (i.e., $M - o(M)$) files. Since $|\mathcal{M}_{\text{↓}}| = M - r + 1$, this condition is equivalent to $M - r = o(M)$.

**LEMMA 1** [$l, r$ AND CONDITIONS FOR ALMOST EMPTY $\mathcal{M}_{\text{↓}}$]: *If $\mathcal{M}_{\text{↓}} \approx \emptyset$, then $r \sim M$, and, moreover,*

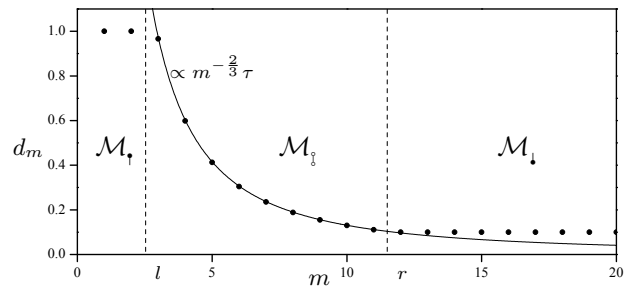


Fig. 1. Density $d_m$, partitions $\mathcal{M}_{\text{↑}}, \mathcal{M}_{\text{⇂}}, \mathcal{M}_{\text{↓}}$, and the $m^{-\frac{2}{3}\tau}$ law (line) of (2b) using the Zipf law of (3).

- *for $\tau < 3/2$, it is*

$$\begin{cases} l \to 1, & \text{if } K \overset{\lim}{<} M^{1-\frac{2\tau}{3}} \\ l \sim \left(\frac{3-2\tau}{3}\right)^{\frac{3}{2\tau}} \frac{K^{\frac{3}{2\tau}}}{M^{\frac{3}{2\tau}-1}}, & \text{if } M^{1-\frac{2\tau}{3}} \overset{\lim}{\leq} K = o(M), \\ l \sim \alpha K, & \text{if } K \sim \beta_\tau M, \end{cases}$$

*where* $\alpha \in (0,1]$, $\beta_{\alpha,\tau} \triangleq \alpha^{\frac{2\tau}{3-2\tau}} \left[\frac{3-2\tau(1-\alpha)}{3}\right]^{\frac{-3}{3-2\tau}}$ *and*

$$\begin{cases} \mathcal{M}_{\downarrow} = \emptyset, & \text{if } \omega(K) = M \overset{\lim}{<} \frac{3-2\tau}{3} KN, \\ \mathcal{M}_{\downarrow} \approx \emptyset, & \text{if } \omega(K) = M \overset{\lim}{\leq} \frac{3-2\tau}{3} KN, \\ \mathcal{M}_{\downarrow} = \emptyset, & \text{if } K = \Theta(M); \end{cases}$$

- *for $\tau = 3/2$, it is*

$$\begin{cases} l \to 1 & \text{if } K \overset{\lim}{\leq} \ln M, \\ l \sim \frac{K}{\ln M} & \text{if } \ln M \overset{\lim}{<} K = o(M), \\ l \sim \alpha K, & \text{if } K \sim \gamma_\alpha M, \end{cases}$$

*where* $\alpha \in (0,1]$, $\gamma_\alpha \triangleq \frac{1}{\alpha} e^{\frac{\alpha-1}{\alpha}}$, *and*

$$\begin{cases} \mathcal{M}_{\downarrow} = \emptyset, & \text{if } M = \omega(K), \text{ and } M \ln M \overset{\lim}{<} KN, \\ \mathcal{M}_{\downarrow} \approx \emptyset, & \text{if } M = \omega(K), \text{ and } M \ln M \overset{\lim}{\leq} KN, \\ \mathcal{M}_{\downarrow} = \emptyset, & \text{if } K = \Theta(M); \end{cases}$$

- *for $\tau > 3/2$, it is*

$$\begin{cases} l \sim \frac{2\tau-3}{2\tau} K, & \text{if } K = o(M), \\ l \sim \alpha K, & \text{if } K \sim \delta_{\alpha,\tau} M, \end{cases}$$

*with* $\alpha \in \left(\frac{2\tau-3}{2\tau}, 1\right]$, $\delta_{\alpha,\tau} \triangleq \alpha^{\frac{2\tau}{3-2\tau}} \left(\frac{3-2\tau(1-\alpha)}{3}\right)^{\frac{-3}{3-2\tau}}$ *and*

$$\begin{cases} \mathcal{M}_{\downarrow} = \emptyset, & \text{if } \omega(K) = M \overset{\lim}{<} \frac{2\tau-3}{2\tau} KN^{\frac{3}{2\tau}}, \\ \mathcal{M}_{\downarrow} \approx \emptyset, & \text{if } \omega(K) = M \overset{\lim}{\leq} \frac{2\tau-3}{2\tau} KN^{\frac{3}{2\tau}}, \\ \mathcal{M}_{\downarrow} = \emptyset, & \text{if } K = \Theta(M). \end{cases}$$

Note that $K - l + 1 = \Theta(K)$ if $K \overset{\lim}{<} M$. If $K \sim M$, then $l \sim K \sim M$, i.e., almost all files are stored locally.

### D. $l$ and $r$ on Non-empty $\mathcal{M}_{\downarrow}$

If $\mathcal{M}_{\downarrow}$ is non-empty, then $M - r = \Theta(M)$.

**LEMMA 2** [$l$ AND $r$ FOR NON-EMPTY $\mathcal{M}_{\downarrow}$]: *If $\mathcal{M}_{\downarrow} \neq \emptyset$, and $KN - M = O(1)$, then $l \to 1$ and $r = \Theta(1)$; in particular,*

$$\begin{cases} r \approx 1 + \frac{3-2\tau}{2\tau}(KN - M), & \text{if } \tau < 3/2, \\ (r-1)\ln(r-1) \approx KN - M, & \text{if } \tau = 3/2, \\ r \approx 1 + \frac{2\tau-3}{2\tau} \frac{KN-M}{N^{1-\frac{3}{2\tau}}} & \text{if } \tau > 3/2. \end{cases}$$

*Else, if $\mathcal{M}_{\downarrow} \neq \emptyset$, and $KN - M = \omega(1)$, then*

- *for $\tau < 3/2$,*
  - *if $KN - M \overset{\lim}{\leq} \frac{2\tau}{3-2\tau} N^{\frac{3}{2\tau}}$, then $l \to 1$,* $r \sim \frac{3-2\tau}{2\tau}(KN - M)$,
  - *if $KN - M \overset{\lim}{>} \frac{2\tau}{3-2\tau} N^{\frac{3}{2\tau}}$, then $l \sim \frac{3-2\tau}{2\tau} \frac{KN-M}{N^{\frac{3}{2\tau}}}$,* $r \sim \frac{3-2\tau}{2\tau}(KN - M)$,
- *for $\tau = 3/2$,*
  - *if $KN - M \overset{\lim}{\leq} N \ln N$, then $l \to 1$,* $r \ln r \sim KN - M$,
  - *if $KN - M \overset{\lim}{>} N \ln N$ then $l \sim \frac{KN-M}{N \ln N}$,* $r \sim \frac{KN-M}{\ln N}$,
- *for $\tau > 3/2$*
  - *if $KN - M \overset{\lim}{\leq} \frac{2\tau}{2\tau-3} N$, then $l \to 1$,* $r \sim \left(\frac{2\tau-3}{2\tau}\right)^{\frac{3}{2\tau}} (KN - M)^{\frac{3}{2\tau}}$,
  - *if $KN - M \overset{\lim}{>} \frac{2\tau}{2\tau-3} N$, then $l \sim \frac{2\tau-3}{2\tau}\left(K - \frac{M}{N}\right)$,* $r \sim \frac{2\tau-3}{2\tau} \frac{KN-M}{N^{1-\frac{3}{2\tau}}}$,

Note, that on all the cases, $K - l + 1 = \Theta(K)$.

### E. Capacity scaling

Now, we are ready to find the asymptotic behavior of the rate $C$ in the various regimes of the size of caches $K$ vs. the number of nodes $N$ vs. the number of files $M$.

First, we establish the Gupta-Kumar rate [2] $O\left(\sqrt{N}\right)$ as an upper bound. This is intuitive: if replication is ineffective (e.g., due to large number of files or small size of caches), then the system and its performance essentially reduce to [2].

**LEMMA 3** [UPPER BOUND ON $C$]: $C = O\left(\sqrt{N}\right)$.

Next, we begin the asymptotic analysis using the results for $l$ and $r$ obtained in Lemmas 1 and 2. In the case of almost empty $\mathcal{M}_{\downarrow}$, as $M - r = o(M)$, there are very few files stored once in the network; therefore, $C$ is determined by $C_{\wr}$.

**THEOREM 4** [CAPACITY FOR ALMOST EMPTY $\mathcal{M}_{\downarrow}$]: *If $K \sim M$, then*

- *if $\tau < \frac{3}{2}$,* $C = \Theta\left(\frac{M-l}{M}\right)$,
- *if $\tau = \frac{3}{2}$,* $C = \Theta\left(\left(\frac{M-l}{M}\right)^{\frac{3}{2}}\right)$,
- *if $\tau > \frac{3}{2}$,* $C = \Theta\left(\frac{M-l}{M^{\tau}}\right)$.

*If $K \overset{\lim}{<} M$, and $\mathcal{M}_{\downarrow} \approx \emptyset$, then*

- *if $\tau < 1$,* $C = \Theta\left(\frac{\sqrt{M}}{\sqrt{K}}\right)$,
- *if $\tau = 1$,* $C = \Theta\left(\frac{\sqrt{M}}{\sqrt{K}\log M}\right)$,
- *if $1 < \tau < \frac{3}{2}$,* $C = \Theta\left(\frac{M^{\frac{3}{2}-\tau}}{\sqrt{K}}\right)$,
- *if $\tau = \frac{3}{2}$, and $K = \Theta(M)$,* $C = \Theta\left(\frac{1}{\sqrt{K}}\right)$,
- *if $\tau = \frac{3}{2}$, and $K = o(M)$,* $C = \Theta\left(\frac{\log^{3/2} M}{\sqrt{K}}\right)$,
- *if $\tau > \frac{3}{2}$,* $C = \Theta\left(\frac{1}{\sqrt{K}}\right)$.

**THEOREM 5** [CAPACITY FOR NON-EMPTY $\mathcal{M}_{\downarrow}$]:

- *If* $\tau<1$, $C=\Theta\left(\sqrt{N}\right)$,
- *if* $\tau=1$, *and* $M \overset{\lim}{<} KN$, $C=\Theta\left(\frac{\sqrt{N}}{\log M}\right)$,
- *if* $\tau=1$, *and* $M \sim KN$, $C=\Theta\left(\sqrt{N}\right)$,
- *if* $1<\tau<\frac{3}{2}$, $C=\Theta\left(\frac{\sqrt{N}}{(KN-M)^{\tau-1}}\right)$,
- *if* $\tau=\frac{3}{2}$, $C=\Theta\left(\frac{\log^{\frac{3}{2}} r}{\sqrt{K-\frac{M}{N}}}\right)$,
- *if* $\tau>\frac{3}{2}$, *and* $KN-M \overset{\lim}{\leq} \frac{2\tau}{2\tau-3}N$, $C=\Theta\left(\frac{\sqrt{N}}{\sqrt{NK-M}}\right)$,
- *if* $\tau>\frac{3}{2}$, *and* $KN-M \overset{\lim}{>} \frac{2\tau}{2\tau-3}N$, $C=\Theta\left(\frac{N^{\tau-1}}{(NK-M)^{\tau-1}}\right)$.

## IV. DISCUSSION ON ASYMPTOTIC LAWS

The main result of the asymptotic laws regards the minimum link rate (of the bottleneck link) required to sustain a constant request rate from each node. As a preliminary comment, the link rates are subject to the information theory, e.g. Shannon's capacity law. Thus, a rate $C$ that scales to infinity should be rather interpreted as the inverse of the sustainable request rate $\lambda$, e.g., a result $C=\Theta(\sqrt{N})$ for $\lambda=1$ is equivalent to $C=\Theta(1)$ for the Gupta-Kumar law of $\lambda=\Theta\left(\frac{1}{\sqrt{N}}\right)$.

Power law parameter $\tau$ sets two phase transition points for the values of 1 and $\frac{3}{2}$, leading to distinct asymptotics: the higher $\tau$, the more uneven the popularity of files, and thus, the more advantageous the caching (i.e. lower rate $C$). As summarized in Table II, $C=O(1)$ on $\tau>\frac{3}{2}$, or, i.e., the wireless network is sustainable (it corresponds to a traffic carrying capacity of $O(1)$ in the Gupta-Kumar setup).In real systems, the Zipf parameter ranges typically from 0.5 [13] to 3 [14] depending on the application: low values are typical in routers, intermediate values in proxies and higher values in mobile applications [15], [16]–see also references therein.

More common are the cases of low and intermediate values of $\tau$ (representative also of the whole file population without any application bias), which flatten the popularity distribution towards the uniform. Replication is less effective, ending up to the $\Theta\left(\sqrt{N}\right)$ law for $\tau<1$, a synonym for the Gupta-Kumar law. When $M \overset{\lim}{\leq} \frac{3-2\tau}{3}KN$, then only a few files are cached once (the condition $\mathcal{M}_{\downarrow}\approx\emptyset$) in which case there is an improvement over [2], see Theorem 4.

Comparing our results to [9], [10], we note the differences due to node cache capacity going to infinity as well. First, when $\mathcal{M}_{\downarrow}\approx\emptyset$, the improvement is significant. More precisely, the term $\frac{1}{\sqrt{K}}$ multiplies the asymptotic law, suggesting that the required link rate of the bottleneck can be partially mitigated by investment in caching. On the other hand, when $\mathcal{M}_{\downarrow}\neq\emptyset$, the term $K$ appears only for $\tau>1$, in the form of $KN-M$. Note, however, that the condition for $\mathcal{M}_{\downarrow}\approx\emptyset$ depends itself on $K$, see Lemma 1. In particular, a sufficient increase in $K$ can guarantee this condition. Thus, if $\tau\leq 1$, investment in cache size makes sense only if it suffices to guarantee $\mathcal{M}_{\downarrow}\approx\emptyset$.

## V. CONCLUSIONS & FUTURE WORK

In this work, we investigated the effect of caching in the asymptotic capacity of wireless networks under the paradigm of content replication and delivery. Depending on the file popularity distribution, there exist regimes of network expansion where caching can be effective tool in mitigating the problem of multihop wireless networks sustainability. Specifically, in the regime of $\mathcal{M}_{\downarrow}\approx\emptyset$, increasing the cache size introduces a $\frac{1}{\sqrt{K}}$ multiplicative term in the required link capacity. Also, in $\mathcal{M}_{\downarrow}\neq\emptyset$, increasing the cache size is helpful if $\tau>1$.

A future extension of this work will focus on establishing the result on non-symmetric topologies and arrival conditions. Also, we are interested in studying the effect of in-network caching in medium-sized wireless networks.


## ACKNOWLEDGMENTS

This work has been supported by the European Commission through the FP7 projects STAMINA 265496 and PURSUIT 257217.


## REFERENCES


[1] Cisco, "*Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 20102015*," White Paper, 2011, Tech. Rep.

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

[3] J. Hennesy and D. Patterson, *Computer Architecture: A Quantitative Approach*. San Francisco, CA, USA: Morgan-Kauffman, 4th ed., 2007.

[4] 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.

[5] 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.

[6] M. Franceschetti, O. Dousse, D. Tse, and P. Thiran, "Closing the gap in the capacity of wireless networks via percolation theory," *IEEE Trans. Inf. Theory*, vol. 53, pp. 1009–1018, Mar. 2007.

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

[8] 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.

[9] S. Gitzenis, G. S. Paschos, and L. Tassiulas, "Asymptotic laws for content replication and delivery in wireless networks," in *Proc. of INFOCOM*, Orlando, FL, USA, Mar. 2011.

[10] ——, "*Asymptotic Laws for Joint Content Replication and Delivery in Wireless Networks*," arXiv:1201.3095v1 [cs.NI], Tech. Rep.

[11] 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.

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

[13] 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.

[14] 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.

[15] 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.

[16] 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.

TABLE II

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

| $M \div K \div N$ | | $K \sim M$ | $K \sim \beta_{\alpha,\tau} M$ | $M^{1-\frac{2\tau}{3}} \overset{\lim}{\leq} K$ | $M^{1-\frac{2\tau}{3}} \overset{\lim}{>} K$ | $M \overset{\lim}{<} \frac{3-2\tau}{3} KN$ | $M \sim \frac{3-2\tau}{3} KN$ | $M \overset{\lim}{>} \frac{3-2\tau}{3} KN$ and $KN-M \overset{\lim}{>} \frac{2\tau}{3-2\tau} N^{\frac{3}{2\tau}}$ | $M \overset{\lim}{>} \frac{3-2\tau}{3} KN$ and $\frac{2\tau}{3-2\tau} N^{\frac{3}{2\tau}} \overset{\lim}{\geq} KN-M$, and $KN-M = \omega(1)$ | $M \overset{\lim}{>} \frac{3-2\tau}{3} KN$ and $KN-M = O(1)$ |
|---|---|---|---|---|---|---|---|---|---|---|
| $l$ | | $\sim K$ | $\sim \alpha K$ | $\sim \left(\frac{3-2\tau}{3}\right)^{\frac{3}{2\tau}} \frac{K^{\frac{3}{2\tau}}}{M^{\frac{3}{2\tau}-1}}$ | $\to 1$ | | | $\sim \frac{3-2\tau}{2\tau} \frac{KN-M}{N^{\frac{3}{2\tau}}}$ | $\to 1$ | |
| $r$ | | $M+1$ | | | | $M+1$ | $\sim M$ | $\sim \left(\frac{3-2\tau}{2\tau}\right)(KN-M)$ | | |
| $\mathcal{M}_{\downarrow}$ | | empty | | | | empty | almost empty | non-empty | | |
| $C$ | $\tau < 1$ | $\Theta\left(\frac{M-l}{M}\right)$ | $\Theta\left(\sqrt{\frac{M}{K}}\right)$ | | | | | $\Theta\left(\sqrt{N}\right)$ | | |
| | $\tau = 1$ | $\Theta\left(\frac{M-l}{M}\right)$ | $\Theta\left(\frac{\sqrt{M}}{\sqrt{K}\log M}\right)$ | | | | | $\Theta\left(\frac{\sqrt{N}}{\log M}\right)$ | | $\Theta\left(\sqrt{N}\right)$ |
| | $1 < \tau < \frac{3}{2}$ | $\Theta\left(\frac{M-l}{M}\right)$ | $\Theta\left(\frac{M^{\frac{3}{2}-\tau}}{\sqrt{K}}\right)$ | | | | | $\Theta\left(\frac{\sqrt{N}}{(KN-M)^{\tau-1}}\right)$ | | |

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

| $M \div K \div N$ | $K \sim M$ | $K \sim \gamma_{\alpha} M$ | $\ln M \overset{\lim}{<} K$ | $\ln M \overset{\lim}{\geq} K$ | $M \ln M \overset{\lim}{<} KN$ | $M \ln M \sim KN$ | $M \ln M \overset{\lim}{>} KN$ and $KN-M \overset{\lim}{>} N \ln N$ | $M \ln M \overset{\lim}{>} KN$ and $N \ln N \overset{\lim}{\geq} KN-M$ |
|---|---|---|---|---|---|---|---|---|
| $l$ | $\sim K$ | $\sim \alpha K$ | $\sim \frac{K}{\ln M}$ | $\to 1$ | | | $\sim \frac{KN-M}{N \ln N}$ | $\to 1$ |
| $r$ | $M+1$ | | | | $M+1$ | $\sim M$ | $\sim \frac{KN-M}{\ln N}$ | $r \ln r \sim KN-M$ |
| $\mathcal{M}_{\downarrow}$ | empty | | | | empty | almost empty | non-empty | |
| $C$ | $\Theta\left(\left(\frac{M-l}{M}\right)^{\frac{3}{2}}\right)$ | $\Theta\left(\frac{1}{\sqrt{K}}\right)$ | $\Theta\left(\frac{\log^{\frac{3}{2}} M}{\sqrt{K}}\right)$ | | | | $\Theta\left(\log^{\frac{3}{2}} r \sqrt{\frac{N}{KN-M}}\right)$ | |

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

| $M \div K \div N$ | $K \sim M$ | $K \sim \delta_{\alpha,\tau} M$ | $K = o(M)$ | $M \overset{\lim}{<} \frac{2\tau-3}{2\tau} KN^{\frac{3}{2\tau}}$ | $M \sim \frac{2\tau-3}{2\tau} KN^{\frac{3}{2\tau}}$ | $M \overset{\lim}{>} \frac{2\tau-3}{2\tau} KN^{\frac{3}{2\tau}} N$ and $KN-M \overset{\lim}{>} \frac{2\tau}{2\tau-3} N$ | $M \overset{\lim}{>} \frac{2\tau-3}{2\tau} KN^{\frac{3}{2\tau}} N$ and $\frac{2\tau}{2\tau-3} N \overset{\lim}{\geq} KN-M$ |
|---|---|---|---|---|---|---|---|
| $l$ | $\sim K$ | $\sim \alpha K$ | $\sim \frac{2\tau-3}{2\tau} K$ | | | $\sim \frac{2\tau-3}{2\tau}\left(K - \frac{M}{N}\right)$ | $\to 1$ |
| $r$ | $M+1$ | | | $M+1$ | $\sim M$ | $\sim \frac{2\tau-3}{2\tau} \frac{KN-M}{N^{1-\frac{3}{2\tau}}}$ | $\sim \left(\frac{2\tau-3}{2\tau}\right)^{\frac{3}{2\tau}} (KN-M)^{\frac{3}{2\tau}}$ |
| $\mathcal{M}_{\downarrow}$ | empty | | | empty | almost empty | non-empty | |
| $C$ | $\Theta\left(\frac{M-l}{M^{\tau}}\right)$ | $\Theta\left(\frac{1}{\sqrt{K}}\right)$ | | | | $\Theta\left(\frac{N^{\tau-1}}{(KN-M)^{\tau-1}}\right)$ | $\Theta\left(\sqrt{\frac{N}{KN-M}}\right)$ |