Enhancing Wireless Networks with Caching: Asymptotic Laws, Sustainability & Trade-offs

S. Gitzenis[a], G. S. Paschos[b,a], L. Tassiulas[c,a]

[a] *Information Technologies Institute, Centre of Research & Technology, Hellas, 6thkm Charilaou-Thermi Rd., 570001, Thessaloniki, Greece*
[b] *Laboratory of Information and Decision Systems, Massachusetts Institute of Technology, 77 Vassar St. Cambridge, MA, USA*
[c] *Dept. of Computer and Communication Engineering, University of Thessaly, 37 Glavani Str., 38221 Volos, Greece*

Abstract

We investigate on the sustainability of multihop wireless communications in the context of Information-Centric Networks, when content is replicated in caches over the network. The problem is cast in a flat wireless network for a given content popularity distribution and sized by three parameters, (i) the network size N, (ii) the content volume M and (iii) the cache capacity K per node. The objective is to select a joint replication and delivery scheme that minimizes the link traffic. Assuming the Zipf distribution about the content popularity, a law well established in the research on Internet traffic, we compute an order optimal solution, let the three size parameters jointly scale to infinity, and find the scaling laws about the link rates, ranging from $O\left(\sqrt{N}\right)$ down to $o(1)$. Analyzing the derived laws, we determine the regimes that the network becomes sustainable subject to the scaling of the three network size parameters and the Zipf rank exponent, characterize the relative merit of network resources and identify the induced trade-offs about network expansion.

Keywords: Scaling laws, wireless network capacity, network sustainability, content centric networks, caching, Zipf Law

1. Introduction

Over the last years, the model of Information Centric Networking (ICN) is receiving increasing attention [1–3]. In this paradigm, user requests are placed on named content basis via host-to-content primitives, as opposed to the address of the node that hosts the desired content using host-to-host primitives. A key motivator for this shift regards the new way content is stored and distributed to the nodes, which enables the seamless replication of the content across the network; large populations of geographically dispersed users can be served at reduced network load and low latency—a major consideration in view of the proliferation of bandwidth-hungry services, such as HD, 3D and multiview video, and P2P services. Content Delivery Networks (CDNs) overlaid in today's Internet already implement replication and reap such benefits.

At the same time, networking is marked with a shift to wireless communications towards supporting user mobility and promoting ubiquitous computing; by 2015, traffic from wireless devices is expected to dominate the total traffic [4]. Despite their extensive adoption, wireless is still mostly confined to single-hop cellular-like deployment at the network edge, a result of the extensive wired telecommunication infrastructure existing in many and major parts of world and the abundant capacity of optical links. However, let alone plain theoretical interest, researchers have always been looking for wireless-only architectures to provide communication when wired infrastructure is absent, as in special scenarios (e.g., Adhoc, Vehicular, or Sensor networks), or for regular Internet access, or even to reassess the economics in comparison to the wired backbone paradigm considering advances such as powerful inexpensive wireless devices, interference combating, cognitive radios or cooperative transmissions.

In this context, the networking community has been investigating on the the sustainability of wireless networks and the emerging scaling laws as they grow in size.In their seminal work [5], Gupta and Kumar studied the asymptotic behavior of planar multihop wireless networks where each node pairs with some other independently selected node to communicate, for a total of N pairs equal to the number of nodes.The per pair data rate decreases as $O(1/\sqrt{N})$, with the denominator accounting for the number hops needed per pair on average. Unfortunately, this law argues against the sustainability of multihop unicast communication motivating, thus, the rethink of wireless networking. In this direction, this work considers the asymptotic behavior under the novel ICN paradigm.

Clearly, caching at the network level can play a key role in reducing the required hops by storing the data close to the clients. The central issue, hence, is the replication, i.e., how much dense, and where content is cached. Along this, routing is a quite important decision. Careless selection

Email addresses: `sgitz@iti.gr` (S. Gitzenis), `gpasxos@gmail.com` (G. S. Paschos), `leandros@uth.gr` (L. Tassiulas)

of the delivery paths may lead to large amounts of traffic traversing the same links multiple times back and forth, wasting network capacity and causing overloads.

In this thread of research, we set out to compute the asymptotic laws of the wireless networks under the *anycast* transport (when data can be retrieved from potentially multiple nodes), and investigate on the sustainability of such networks, in the spirit of [5] and other works. In our forerunner study [6], we carried out an important part of this work which regards the modeling and formulation of the joint problem of replication and delivery. This is a highly complex, and, therefore, intractable optimization that deals with the contents of every cache and delivery paths of every pair of content and network node. Fortunately, it reduces to a simple replication problem whose optimal solution is of the same order with the original—Section 2 summarizes briefly these results.

Using the optimal in-order solution of the problem, we focus on the derivation of the laws and the sustainability issue. Although a set of asymptotic laws has been already derived for the two dimensional space of the network size N and content volume M scaling jointly to infinity (assuming the Zipf distribution about content popularity) in [6, 7], this investigation was incomplete in the sense that networks scale up in their nodes and the hosted content volume, but in the node cache size K, too. In plain words, not only are new nodes added, but the existing ones are upgraded as storage gets abundant and inexpensive.[1]

To complete the set of asymptotic laws and give a comprehensive answer to the sustainability question, this study[2] extends the previous investigation adding the third scaling dimension of the node caching capacity K. In Section 3, we identify all the possible regimes that parameters K, N and M can jointly scale to infinity, and derive closed form expressions about the scaling of the link load.

Next, we focus on the main questions, whether caching leads to better scaling than the unicast, and, in particular, if it can turn wireless networking sustainable. Section 4 provides a comprehensive analysis regarding (i) the precise characterization of the scaling regimes that turn the network sustainable, (ii) the evaluation of the relative merit of network resources in terms of increasing the number of nodes vs. the individual cache capacity, and (iii) the identification of the associated trade-offs. An in-depth presentation of the derived scaling laws is provided in Appendix Appendix A for further probing.

Last, Section 5 recapitulates this study and discusses extensions and future research directions, including relaxing the symmetry assumptions taken here.

1.1. Related Work

In the area of the asymptotically characterizing wireless networking, [5] spurred a series of works often aspiring to overrule the $O(1/\sqrt{N})$ law, applying various traffic models/services and topologies, such as multicast, many-to-one [10], hybrid adhoc with cellular-like infrastructure support [11]. Departing from the conventional multihop communications, [12] considered the novel paradigm of cooperative transmissions over long links. However, the $O(1/\sqrt{N})$ bound was shown to arise from geometry considerations [13], hence it is not possible to breach. Other efforts exploit node mobility leading to a novel paradigm where packets propagate through the physical movement of the carrying node in addition to wireless transmissions, e.g. [14, 15].

On the other hand, the technique of caching has been successfully applied in various domains of computing and networking. In ICNs, the joint optimization of the contents of all caches in an arbitrary network of asymmetric traffic is considered 'daunting' [16]. In contrast, in planar wireless networks with symmetric user requests, it is possible to compute an in-order optimal allocation, as in [6, 17, 18].

In particular, [17] assumes a model of nodes placed randomly and uniformly, as opposed to the regular grid of this and prior works [6, 7]. This results to a problem quite similar to ours, but with some important differences, which would yield different asymptotics had the authors derived the associated scaling laws. On the other hand, [18] considers randomly placed and mobile nodes and investigates on how fast nodes can move before performance is affected. Finally, all [6, 7, 18] ignore the cache size scaling and its ramifications to the sustainability issue.

In a different direction than this and the above, [19] investigates on data delivery using cooperative transmissions [19] in the spirit of [12]; this leads to a hierarchical tree structure of transmissions over arbitrarily long links, as opposed to the short links and shortest path delivery in the multihop communication paradigm. Equally important, [19] does not optimize the replication; cache contents are given, so the optimization is only about the delivery. Last, an arbitrary traffic matrix is assumed, leading to capacity regions, which is more general, hence, stronger than our symmetric approach. However, such results are not as practical to probe on the issue of network sustainability, as capacity regions are complex objects in comparison to simple, closed-form expressions, e.g. the $\Theta\left(\sqrt{M/K}\right)$ law.

2. Model, Problem Formulation & Solution

Here and in Section 3.1, we provide a brief overview of the network model, the optimization problem and its solution, summarizing the pertinent results of [6].

2.1. Networking Model

Consider a network of N identical nodes arranged in a square grid lattice. Nodes are indexed by $n \in \mathcal{N} \triangleq \{1, 2, \ldots, N\}$. Each one is connected to the four neighbors adjacent on the same row or column with non-directed links. A scaling 2D network emerges by keeping the node density fixed and increasing the network size N, as in [5].

[1] During the last decades, the areal density of hard disks has gone through periods of doubling per year to doubling every three years. Similarly, DRAM capacity quadruples every three years [8].

[2] Part of this work appeared in the WiOpt 2012 conference [9].

Unlike [5] and other studies, the grid is not a random topology; it has, however, been used in the past [20] for studying the capacity of wireless networks. In particular, our model abstracts away the operations at the PHY and MAC layers, assuming interference-free connectivity to the four immediate neighbors, circumventing thus the associated complexity. Nonetheless, it does capture the essential characteristics of the wireless networks, that is

1. long wireless links are of low capacity; short-range multihop communications is the way to maximize capacity [5], which justifies the immediate neighborhood connectivity of the grid;
2. the network diameter scales as $\sqrt{N}$ in networks with uniformly distributed node placement as in [5]; in contrast, in wireline networks, the diameter scales as $\log N$, as node connectivity follows power laws [21].

The latter is obvious for the grid, leading directly to the $\Theta(1/\sqrt{N})$ law of [5] for the random pair traffic. In fact, this kind of traffic essentially arises if, due to limited cache capacity, data is stored uniquely in the network; our results match the $\Theta(1/\sqrt{N})$ law, validating the suitability of the grid model. A second as important validation comes from [18], which, using random uniform node placement verifies the laws derived in [6, 7] based on the grid topology.

2.2. Content, Requests & Delivery

Nodes (or users therein) place requests to access content indexed by $m \in \mathcal{M} \triangleq \{1, 2, \dots, M\}$. Each node n is equipped with a cache/buffer, whose contents are denoted by $\mathcal{B}_n$, a subset of $\mathcal{M}$. If node n's request regards a file/data m that lies in $\mathcal{B}_n$, then it is served internally. Due to the limited cache capacity, m will typically not be available locally, thus, node n will have to retrieve it from an other node w that keeps m in its cache. Thus for each (n, m) pair, a route (or set of routes) $\mathcal{R}_{n,m}$ should be specified from n to retrieve m using unicast delivery.

Let K be the storage capacity of nodes' caches, measured in the number of files it can store. This means that all files have the same size, placing a constraint on the cardinality of cache contents $|\mathcal{B}_n| \leq K$. Variable sizes can be captured in this framework by splitting each large file into multiple units, and then treating each one independently.

To be a non-trivial and feasible problem, it should be

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

The first inequality implies that each node has to select the files of its cache, while the second requires the network to have sufficient capacity to store all files at least once.

Last, let nodes $n \in \mathcal{N}$ generate requests for data at rate $\lambda = 1$, common to all nodes. Each request regards a particular file $m \in \mathcal{M}$, depending on the file's *popularity* p_m. In essence, p_m is the probability that a request is directed to file m. A key point to observe is that the requests are assumed independent to each other both spatially and temporally, i.e., given any set of past requests at the network does not affect future requests at any nodes.

In this work, we do not consider the case that the users have unequal request rates and diverse content popularity distributions, e.g., according to user classes, or their geographical location. Our goal is to derive closed form expressions that provide a practical characterization of the network sustainability, similar to the $O(1/\sqrt{N})$ law. Asymmetric formulations fall beyond the scope of this work, as they will, at best, further parametrize the asymptotic laws, or produce multi-dimensional capacity regions that would obfuscate the key issue of sustainability. They are, however, interesting directions for follow-up work.

Moreover, we assume that $[p_m]$ does not vary with time. This allows us to seek static solutions about buffer allocation $[B_n]$ and routes $[R_{n,m}]$. Indeed, a time-varying solution would call for a constant in time number of replicas for each file to be optimal, as dictated by the solution (3); moving these replicas around the grid cannot offer any performance advantage. Clearly, the replication is to be decided based on the popularity: popular files should be replicated densely to minimize network traffic.

2.3. Zipf Content Popularity

Although the formulation presented next and the solution (3) applies to any distribution $[p_m]$, in Section 3, we consider the Zipf law, towards deriving scaling expressions in closed-form. Zipf law is defined as follows:

$$p_m = \frac{1}{H_\tau(M)}\, m^{-\tau}, \tag{2}$$

where τ is the parameter of the distribution, adjusting the rate of popularity decline with m; $H_\tau(n) \triangleq \sum_{j=1}^{n} j^{-\tau}$ is the truncated zeta function evaluated at τ (also called the n^{th} *τ-order generalized harmonic number*). An important property of function H for the computation of the asymptotic laws that follows is that the limit $H_\tau \triangleq \lim_{n\to\infty} H_\tau(n)$ is the Riemann zeta function, which converges when $\tau > 1$.

Zipf law has been observed to model well, among others, the content popularity of the traffic of WWW and other types of services in numerous traces in the literature. It has since been adopted as the dominant model in theoretical studies, e.g., [18, 22–27], in both wired and wireless networks, and especially pertaining to caching.

Regarding parameter τ, values lower or close to 1 are reported in most cases; e.g. in web traffic, literature reports 0.986 in [28], 0.64–0.83 in [29], and close to 1 in [30]. Interestingly, [31] discriminates between traffic measured at proxies reporting a value of about 1 versus 1.4–1.6 for traffic measured at a 'busy' web server. Analyzing traces from mobile browsing, [32] reports values in the range of 1–1.5. About User Generated Content (UGC) video, [33] fits popularity with a combination of Zipf and Exponential Cut-off; regarding the Zipf component, the data suggest values of τ in the ranges of about 0.98–1.47 and 0.45–1.09

in two different popular UGC websites. Last, content popularity in P2P systems was fitted with a $\tau = 0.95$ in [34], and in a Video on Demand system with 0.70 in [35].

As there is no conclusive answer about the value of τ, we consider all possibilities in our investigation.

2.4. General Replication-Delivery Problem

Given any $[\mathcal{B}_n]$ replication and delivery routes $[\mathcal{R}_{n,m}]$, it is easy to express the traffic load C_ℓ at any link ℓ. The problem, then, regards minimizing load C_ℓ through suitable choice of $[\mathcal{B}_n]$ and $[\mathcal{R}_{n,m}]$ given the constraints of (i) cache capacity and (ii) storage of each file at least once. This is a joint replication-delivery optimization, which is detailed in [6] as the minimization on the traffic of the mostly loaded link, $\min_{[\mathcal{B}_n],[\mathcal{R}_{n,m}]} \max_\ell C_\ell$. The resulting value is the minimum link *capacity* required for the network to be stable without having to reject data requests. This is the sole performance metric we consider—[18] also studies the user-perceived delay for constant K.

This joint optimization turns out to be a hard combinatorial problem, not amenable to an easy-to-compute solution. Therefore, we resort to simplifications and approximations towards suboptimal but efficient solutions. For our needs, this translates to an order-optimal solution, i.e., whose value of the objective function is within a constant to the optimal, hard-to-compute $\min \max C_\ell$.

A first step that preserves the order-optimality of the solution is the relaxation of the target to the average link traffic $\operatorname{avg}_\ell C_\ell$; then shortest path routes $[\mathcal{R}_{n,m}]$ are optimal [6]. However, the decisive step involves breaking the entanglement among the network caches $[\mathcal{B}_n]$ as seen next.

2.5. 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 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 $m \in \mathcal{M}$, $1/N \leq d_m \leq 1$,
2. $\sum_{m \in \mathcal{M}} d_m \leq K$.

In the above, we optimize on the densities d_m which express the fraction of caches containing file m. In the objective, $d_m^{-\frac{1}{2}} - 1$ approximates (in-order) 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.

Problem 1 is a relaxed version of the original min-max problem [6]. The solutions of both optimizations produce a capacity of the same order. In particular, [6] presents an algorithm that assigns cache contents $\mathcal{B}_n$ from the densities d_m filling the caches in diagonals, and uses shortest path routes for $\mathcal{R}_{n,m}$. As both problems share the same asymptotic laws, it suffices to study the scaling of C.

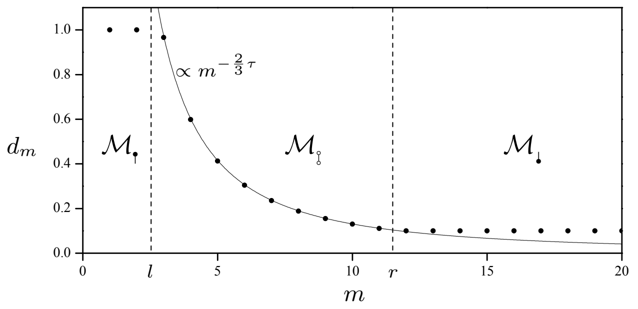


Figure 1: Density d_m, partitions $\mathcal{M}_{\uparrow}, \mathcal{M}_{\updownarrow}, \mathcal{M}_{\downarrow}$, and the $m^{-\frac{2}{3}\tau}$ law (line) of (3b), for the case of content popularity distributed as Zipf.

Note that the optimization of [17] is similar except for the $1/N \leq d_m \leq 1$ bounds. These constraints turn out to affect the solution, and predominantly the asymptotics.

As detailed in [6], the density problem admits a unique solution using the Karush-Kuhn-Tucker (KKT) conditions. Regarding the constraints on d_m about its minimum and maximum value, either one of them can be an equality, or none. This partitions $\mathcal{M}$ into three subsets, the 'up-truncated' $\mathcal{M}_{\uparrow} = \{m : d_m = 1\}$ containing files stored at all nodes, the 'down-truncated' $\mathcal{M}_{\downarrow} = \{m : d_m = 1/N\}$ containing files stored in just one node, and the complementary 'non-truncated' $\mathcal{M}_{\updownarrow} = \mathcal{M} \setminus (\mathcal{M}_{\uparrow} \cup \mathcal{M}_{\downarrow})$ of files with $1/N < d_m < 1$. Arranging p_m in decreasing order, the partitions become $\mathcal{M}_{\uparrow} = \{1, 2, \ldots, l-1\}$, $\mathcal{M}_{\updownarrow} = \{l, l+1, \ldots, r-1\}$, and $\mathcal{M}_{\downarrow} = \{r, r+2, \ldots, M\}$; l and r are integers with $1 \leq l \leq r \leq M+1$.

The solution d_m (see Fig. 1) is equal to

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

3. Asymptotic Laws for Zipf Popularity

Plugging (3) in C, C can be split in three terms

$$C = C_{\updownarrow} + C_{\downarrow} - \sum_{j=l}^{M} p_m,$$

with $\sum_{j=l}^{M} p_m = O(1)$—please see the definition of the asymptotic notation in Table 1. Terms $C_{\updownarrow}$ and $C_{\downarrow}$ express the traffic generated from requests on the files of non-truncated $\mathcal{M}_{\updownarrow}$ and downtruncated $\mathcal{M}_{\downarrow}$, respectively. The latter are files uniquely stored across the network,

Table 1
Definition of asymptotic notation
(f and g are positive functions)

$f=o(g)$	For any $k>0$,	there exists $\hat{x}$ such that $x \geq \hat{x} \Rightarrow f(x) \leq kg(x)$
$f=O(g)$	For some $k>0$	
$f \overset{\lim}{\leq} kg$	For the specified k,	
$f \overset{\lim}{<} k'g$	For any $0<k<k'$,	
$f=\omega(g)$	For any $k>0$,	there exists $\hat{x}$ such that $x \geq \hat{x} \Rightarrow f(x) \geq kg(x)$
$f=\Omega(g)$	For some $k>0$	
$f \overset{\lim}{\geq} kg$	For the specified k,	
$f \overset{\lim}{>} k'g$	For any $k>k'$	
$f=\Theta(g)$	If $f=O(g)$ and $f=\Omega(g)$	
$f \sim g$	As $x \to \infty$, $\frac{f(x)}{g(x)} \to 1$	

hence they lie on average $\Theta\left(\sqrt{N}\right)$ hops away. Therefore, a $\Theta\left(\sqrt{N}\right)$ law for C is avoided (such law matches [5] as explained in Section 4), if the cardinality of the downtruncated $\mathcal{M}_{\downarrow}$ is small. For the Zipf law, $C_{\updownarrow}$ and $C_{\downarrow}$ become

$$C_{\updownarrow} \triangleq \sum_{m \in \mathcal{M}_{\updownarrow}} \frac{p_m}{\sqrt{d_m}} \overset{(2)}{=} \frac{\left[H_{\frac{2\tau}{3}}(r-1) - H_{\frac{2\tau}{3}}(l-1)\right]^{\frac{3}{2}}}{\sqrt{K-l+1-\frac{M-r+1}{N}}\ H_\tau(M)}, \quad (4)$$

$$C_{\downarrow} \triangleq \sum_{m \in \mathcal{M}_{\downarrow}} \frac{p_m}{\sqrt{d_m}} \overset{(2)}{=} \sqrt{N}\ \frac{H_\tau(M) - H_\tau(r-1)}{H_\tau(M)}. \quad (5)$$

3.1. Approximations on $H_\tau(n)$ & Estimates of l and r

Let us bound the $H_\tau(n)$ sum: for $n \geq m \geq 0$,

$$\int_m^n (x+1)^{-\tau} dx \leq H_\tau(n) - H_\tau(m) \leq 1 + \int_{m+1}^n x^{-\tau} dx, \Rightarrow$$
$$\begin{cases} \frac{(n+1)^{1-\tau}-(m+1)^{1-\tau}}{1-\tau} \leq H_\tau(n) - H_\tau(m) \leq \frac{n^{1-\tau}-(m+1)^{1-\tau}}{1-\tau}+1, & \text{if } \tau \neq 1, \\ \ln\frac{n+1}{m+1} \leq H_\tau(n) - H_\tau(m) \leq \ln\frac{n+1}{m+2}, & \text{if } \tau = 1. \end{cases} \quad (6)$$

Next, we derive a system of equations to estimate l and r, and finally the asymptotics of $C_{\updownarrow}$ and $C_{\downarrow}$.

3.1.1. Estimation of l

If l is a valid file index (i.e. the non-truncated set $\mathcal{M}_{\updownarrow}$ and downtruncated $\mathcal{M}_{\downarrow}$ are not both empty), it is $d_l < 1$ which through (3b) expands to

$$K-l+1-\frac{M-r+1}{N} < l^{\frac{2\tau}{3}} \left[H_{\frac{2\tau}{3}}(r-1) - H_{\frac{2\tau}{3}}(l-1)\right]. \quad (7)$$

If, furthermore, $l > 1$ (i.e., the up-truncated set $\mathcal{M}_{\uparrow}$ is not empty), then $d_{l-1} = 1$. The fact that the solution of indexes (l,r) is optimal and valid means that if we attempted to decrease index l by 1, the evaluation of (3b) should yield a density d_{l-1} greater than 1,

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

This is easy to verify: if (8) had the opposite sign, the new solution $(l-1,r)$ would be valid and yield a lower value of C, which contradicts the optimality of the (l,r) solution.

Inequalities (7)-(8) permit computing the integer index l. However, l can be estimated by treating (7) as an approximate equality (given that $d_{l-1} = 1$ and $d_l < 1$):

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

Although (9) does not produce necessarily an integer—there can be a fractional error from the actual integer value, it suffices for the needs of the asymptotic analysis.

3.1.2. Estimation of r

Likewise, if $r \neq 1$ (i.e., $\mathcal{M}_{\updownarrow}$ and $\mathcal{M}_{\downarrow}$ are not both empty), it is $d_{r-1} > \frac{1}{N}$, or equivalently

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

If, moreover, $r \leq M$ (i.e., the down-truncated $\mathcal{M}_{\downarrow}$ is not empty), then $d_r = N^{-1}$. Similarly, r cannot be increased by one, without (3b) producing a density less than $1/N$:

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

Treating, hence, (10) as an approximate equality, we get a non-integer approximation of r

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

3.1.3. Estimation of l/r

When l and r are not equal to the extremes, i.e., $1 < l \leq r < M+1$, it is $d_{l-1}/d_r = N$, or using (3),

$$l \cong r N^{-\frac{3}{2\tau}}. \quad (13)$$

3.2. l and r on Almost Empty $\mathcal{M}_{\downarrow}$

The first case of interest is when the solution's downtruncated set $\mathcal{M}_{\downarrow}$ is almost empty. We define $\mathcal{M}_{\downarrow} \approx \emptyset$ if $|\mathcal{M}_{\downarrow}| = o(M)$, i.e., the number of singly replicated files over their total number is negligible ($\mathcal{M}_{\downarrow} = \emptyset$ is just a special case). To have $\mathcal{M}_{\downarrow} \approx \emptyset$, 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 $|\mathcal{M}_{\downarrow}| = M-r+1$, this is equivalent to $M-r = o(M)$.

Lemma 1 (Almost Empty $\mathcal{M}_{\downarrow}$ Conditions and indices)**.** *If $\mathcal{M}_{\downarrow} \approx \emptyset$, then $r \sim M$, and, moreover,*

- *for* $\tau < 3/2$*, it is*
$$\begin{cases} l \to 1, & \textit{if } K \overset{\lim}{<} M^{1-\frac{2\tau}{3}} \\ l \sim \left(\frac{3-2\tau}{3}\right)^{\frac{3}{2\tau}} \frac{K^{\frac{3}{2\tau}}}{M^{\frac{3}{2\tau}-1}}, & \textit{if } M^{1-\frac{2\tau}{3}} \overset{\lim}{\leq} K = o(M), \\ l \sim \alpha K, & \textit{if } K \sim \beta_{\alpha,\tau} M, \\ M-l \sim \sqrt{\frac{3}{\tau}(M-K)}, & \textit{if } K \sim M, M-K = \omega(1), \end{cases}$$
where $\alpha \in (0,1)$, $\beta_{\alpha,\tau} \triangleq \alpha^{\frac{2\tau}{3-2\tau}} \left[\frac{3}{3-2\tau(1-\alpha)}\right]^{\frac{3}{3-2\tau}}$ *and*
$$\begin{cases} \mathcal{M}_{\downarrow} = \emptyset, & \textit{if } K \overset{\lim}{\leq} M \overset{\lim}{<} \frac{3-2\tau}{3} KN, \\ \mathcal{M}_{\downarrow} \approx \emptyset, & \textit{if } K \overset{\lim}{\leq} M \overset{\lim}{\leq} \frac{3-2\tau}{3} KN, \\ \mathcal{M}_{\downarrow} = \emptyset, & \textit{if } K = \Theta(M); \end{cases}$$
- *for* $\tau = 3/2$*, it is*
$$\begin{cases} l \to 1 & \textit{if } K \overset{\lim}{\leq} \ln M, \\ l \sim \frac{K}{\ln M} & \textit{if } \ln M \overset{\lim}{<} K = o(M), \\ l \sim \alpha K, & \textit{if } K \sim \gamma_\alpha M, \\ M-l \sim M\sqrt{2\frac{M-K}{M}}, & \textit{if } K \sim M, M-K = \omega(1), \end{cases}$$
where $\alpha \in (0,1)$, $\gamma_\alpha \triangleq \frac{1}{\alpha} e^{\frac{\alpha-1}{\alpha}}$*, and*
$$\begin{cases} \mathcal{M}_{\downarrow} = \emptyset, & \textit{if } K = \Theta(M); \\ \mathcal{M}_{\downarrow} = \emptyset, & \textit{if } K = O(M), \textit{ and } M \ln M \overset{\lim}{<} KN, \\ \mathcal{M}_{\downarrow} \approx \emptyset, & \textit{if } K = O(M), \textit{ and } M \ln M \overset{\lim}{\leq} KN, \end{cases}$$
- *for* $\tau > 3/2$*, it is*
$$\begin{cases} l \sim \frac{2\tau-3}{2\tau} K, & \textit{if } K = o(M), \\ l \sim \alpha K, & \textit{if } K \sim \delta_{\alpha,\tau} M, \\ M-l \sim M\sqrt{\frac{3}{\tau}(M-K)}, & \textit{if } K \sim M, M-K = \omega(1), \end{cases}$$
with $\alpha \in \left(1-\frac{3}{2\tau}, 1\right)$, $\delta_{\alpha,\tau} \triangleq \alpha^{\frac{2\tau}{3-2\tau}} \left[\frac{3-2\tau(1-\alpha)}{3}\right]^{\frac{3}{2\tau-3}}$ *and*
$$\begin{cases} \mathcal{M}_{\downarrow} = \emptyset, & \textit{if } K = O(M), \textit{ and } M \overset{\lim}{<} \frac{2\tau-3}{2\tau} KN^{\frac{3}{2\tau}}, \\ \mathcal{M}_{\downarrow} \approx \emptyset, & \textit{if } K = O(M), \textit{ and } M \overset{\lim}{\leq} \frac{2\tau-3}{2\tau} KN^{\frac{3}{2\tau}}, \\ \mathcal{M}_{\downarrow} = \emptyset, & \textit{if } K = \Theta(M). \end{cases}$$

The proof of this and next results is in Appendix B. In the above, we specify the conditions for both the strict case of $\mathcal{M}_{\downarrow} = \emptyset$ and its relaxed variant $\mathcal{M}_{\downarrow} \cong \emptyset$. Note that if $K \sim M$, then $l \sim K \sim M$, i.e., almost all files are stored locally as intuitively expected.

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

If $\mathcal{M}_{\downarrow}$ is non-empty, then $M - r = \Theta(M)$. In practice, this means that M increases fast w.r.t. K and N to avoid d_m dropping below $1/N$ for a non-negligible number of files.

Lemma 2 (Indices for non Almost-Empty $\mathcal{M}_{\downarrow}$)**.** *If* $KN - M = O(1)$*,then* $l \to 1$ *and* $r = \Theta(1)$*; in particular,*
$$\begin{cases} r \approx 1 + \frac{3-2\tau}{2\tau}(KN-M), & \textit{if } \tau < 3/2, \\ (r-1)\ln(r-1) \approx KN-M, & \textit{if } \tau = 3/2, \\ r \approx 1 + \frac{2\tau-3}{2\tau} \frac{KN-M}{N^{1-\frac{3}{2\tau}}} & \textit{if } \tau > 3/2. \end{cases}$$
Else, if $\mathcal{M}_{\downarrow} \not\approx \emptyset$ *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 $l \to 1$*,* $\quad r \sim \frac{3-2\tau}{2\tau}(KN-M)$*,*
  - *if* $KN - M \overset{\lim}{>} \frac{2\tau}{3-2\tau} N^{\frac{3}{2\tau}}$*,*
    then $l \sim \frac{3-2\tau}{2\tau} \frac{KN-M}{N^{\frac{3}{2\tau}}}$*,* $\quad r \sim \frac{3-2\tau}{2\tau}(KN-M)$*,*
- *for* $\tau = 3/2$*,*
  - *if* $KN - M \overset{\lim}{\leq} N \ln N$*,*
    then $l \to 1$*,* $\quad r \ln r \sim KN - M$*,*
  - *if* $KN - M \overset{\lim}{>} N \ln N$
    then $l \sim \frac{KN-M}{N \ln N}$*,* $\quad r \sim \frac{KN-M}{\ln N}$*,*
- *for* $\tau > 3/2$
  - *if* $KN - M \overset{\lim}{\leq} \frac{2\tau}{2\tau-3} N$*,*
    then $l \to 1$*,* $\quad r \sim \left(\frac{2\tau-3}{2\tau}\right)^{\frac{3}{2\tau}} (KN-M)^{\frac{3}{2\tau}}$*,*
  - *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, $K - l + 1 = \Theta(K)$. Observe, moreover, that the asymptotic laws of r for $KN - M = \omega(1)$ agree with the approximations for $KN - M = O(1)$.

3.4. Capacity Scaling

Now, we are ready to compute 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 $O\left(\sqrt{N}\right)$ as an upper bound [5]. This is intuitive: if replication is ineffective (e.g., due to large number of files or small size of caches), then the system essentially reduces to [5].

Lemma 3 (Upper Bound on C)**.** $C = O\left(\sqrt{N}\right)$*.*

Next, we proceed to the asymptotic analysis of C using the results for l and r from Lemmas 1 and 2.

Theorem 4 (Capacity for Almost Empty $\mathcal{M}_{\downarrow}$)**.**
If $K \sim M$*, then*

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

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

- *if* $\tau < 1$*,* $\quad C = \Theta\left(\sqrt{\frac{M}{K}}\right)$*,*
- *if* $\tau = 1$*,* $\quad C = \Theta\left(\frac{1}{\log M}\sqrt{\frac{M}{K}}\right)$*,*

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

Theorem 5 (Capacity for Non-Empty $\mathcal{M}_{\downarrow}$)**.** *If* $|\mathcal{M}_{\downarrow}| = \Omega(M)$*, then*

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

4. Scaling Laws, Sustainability & Trade-offs

In this section, we discuss the scaling laws and their effects in the sustainability of evolving networks. The first result of our study is the rate C of the most loaded link, which is computed under the assumption that each node places requests with rate $\lambda = 1$. This sets the lower bound on the link capacity so that the network can sustain the offered traffic. Moreover, C also expresses order-wise the average number of hops required to serve a user request.

The attainable wireless link capacity is subject to Information Theory fundamentals and available wireless communication technology. That is, increases in link capacity are expected to come from the use of additional frequency bands, and advanced communication techniques, such as MIMO. These advances, however, can often be costly and technically challenging to realize. Indeed, in the context of the random communicating pairs of [5], such advances in link capacity were ignored; C was assumed constant as the network size N scales up, therefore the rate λ of communication per flow declines as $1/\sqrt{N}$.

Here, likewise, we take the perspective that link rates cannot increase (especially as fast as $\sqrt{N}$), if we wish to keep λ constant. Therefore, when studying the scaling laws of C, central is the question of how to keep it *bounded*, i.e. $O(1)$, by *appropriately matching parameters K, N and M* so as to ensure the sustainability of the network.

As a first comment, in our study, the law $C = \Theta\left(\sqrt{N}\right)$ for $\lambda = 1$ arises on low replication capacity$KN - M = \Theta(1)$, i.e., a negligible number of files can be stored in multiple nodes.In a practical system, this would be recast as $C = \Theta(1)$ for $\lambda = \Theta({}^{1}/\sqrt{N})$, which restates [5].

4.1. Asymptotic Laws Systemization and Analysis

Lemmas 1–2 and Theorems 4–5 provide the complete set of laws regarding the asymptotic behavior of C, l, r and singly replicated set of files $\mathcal{M}_{\downarrow}$ in a minimal mathematically complete form; unfortunately, this is not convenient to probe into the behavior of the system given the 3D space of the scaling parameters K, N and M. Appendix A organizes and analyzes in depth the laws and provides equivalent expressions and bounds, along with visual aids, that facilitate the reader understand the findings of the last Section and interpret the order of the derived laws through Figure A.

Given the Riemman-Zeta function H_τ phase transition at $\tau = 1$ and that it appears parametrized on τ and ${}^{2\tau}/_{3}$ in (4)-(5), the solution exhibits two phase transition points on τ for the values of 1 and ${}^{3}/_{2}$, resulting in distinct scaling laws, as reflected in the preceding theorems and Figure A.

Next, we focus on the key issue of system sustainability. To assist this search, we compile a high-level synopsis of the asymptotic laws of C into Table 2.

4.2. The Role of the Zipf Parameter τ in the Scaling of C

As depicted in Table 2, the system behavior varies much with τ and the scaling laws take much diverse forms. Specifically, the higher τ, the lower the order of C becomes. This is, for example, evident for the important case of $C = O(1)$: the conditions are much easier to satisfy for $\tau > {}^{3}/_{2}$ than $\tau < {}^{3}/_{2}$. Moreover, the higher τ from ${}^{3}/_{2}$, the more relaxed the condition becomes. On the other hand, on $\tau < 1$, $C = O(1)$ requires that cache size K is a fraction (same order) of the content volume M, an undoubtedly quite strict condition.

This behavior on τ is explained intuitively by the popularity distribution. The higher τ, the more uneven the popularity of files is; in other words, for high τ, low index m popular files become increasingly more popular in comparison to the unpopular high index m ones.

In fact, it is this asymmetry that makes caching efficient. For $\tau > {}^{3}/_{2}$, it is not necessary that the cache size K is a fraction of the content volume M to keep $C = O(1)$. The sparsely replicated files, distant to access on average, are too unpopular to affect C; it is the densely stored files, accessible at a minimal load that determine the (low) order of C. In contrast, as τ approaches 0, the distribution flattens toward uniform; then, all files become equally popular, to be replicated, therefore, with the same density. This explains why cache size K should be a fraction of content volume M to keep the required capacity C bounded.

Last, another important point to observe is that C decreases with τ. This persists in all cases of Table 2 as well as all the detailed laws, listed in Figure A of Appendix A.

Table 2
A synopsis of the scaling laws.
Laws for $\tau = 1$ and $\tau = 3/2$ are similar to the laws of $\tau < 1$ and $\tau > 3/2$, respectively.

τ	'Low' Link Capacity		'High' Link Capacity	
	Condition	C	Condition	C
$\tau < 1$	$K \sim M$ or $\frac{3-2\tau}{3}KN \overset{\lim}{\geq} M$	$O\left(\sqrt{\frac{M}{K}}\right)$	$\frac{3-2\tau}{3}KN \overset{\lim}{<} M$	$\Theta\left(\sqrt{N}\right)$
$1 < \tau < 3/2$	$K \sim M$ or $\frac{3-2\tau}{3}KN \overset{\lim}{\geq} M$	$O\left(\frac{M^{\frac{3}{2}-\tau}}{\sqrt{K}}\right)$	$\frac{3-2\tau}{3}KN \overset{\lim}{<} M$	$\Theta\left(\frac{\sqrt{N}}{(KN-M)^{\tau-1}}\right)$
$\tau > 3/2$	$K \sim M$ or $\frac{2\tau-3}{2\tau}KN^{\frac{3}{2\tau}} \overset{\lim}{\geq} M$	$O(1)$	$\frac{2\tau-3}{2\tau}KN^{\frac{3}{2\tau}} \overset{\lim}{<} M$	$\Theta\left(\sqrt{\frac{N}{KN-M}}\right)$

4.3. The Role of Parameters K, N, M in Evolving Networks

Let us consider networks that grow in time as regards K, N and M as part of a network expansion. Although the expansion over time is not the only possibility of network evolution, it nevertheless constitutes an intuitive example.

As a first example, we can consider a network centrally managed that has to cater for its user needs. In an evolving network, the content volume M is expected to increase both spontaneously, as the existing users inject more content in it or demand new content from it, as well as due to network expansion. The latter involves increasing the node count N. The users at the new nodes are expected to bring in their own content, as well as demand new content from the network to join; both of them result in increasing M. Hence, with respect to the existing users' content generation, M is an exogenous parameter; however, taking into account network expansion, M can have an additional component, which, as discussed next is controllable.

Regarding the node cache capacity K, it is fair to treat it as a parameter in control of the network operator. K should be such to, at least, maintain storage for all data in the view of $M \leq KN$ constraint of (1); in reality, it should be even higher in order to preserve the desired Quality of Service, a consideration related to sustainability. As for the left side of (1), K will not be as large as M; i.e., if each node possessed all content, there is no need for communication among nodes in the setup of static content studied here (see Section 5 about time-varying content).

As for network size, N is a parameter that can be considered controllable (e.g., in a cellular setup, the operator takes decisions on network expansion along with upgrades in node cache size K). As aforesaid, K and N are linked against M through the need to provide sufficient storage to maintain the primary copy of the content.

In a second example of an adhoc network (in contrast to the previous centralized one), the network is run by a community of users who spontaneously join and increase both content volume M and node count N. Moreover, they control their own nodes' cache capacity K. Provided that the nodes are sufficiently uniformly positioned (see discussion in Section 5 about deviations), the asymptotic laws of the grid model are applicable.

In both of the above paradigms, the key question revolves about the sustainability of the network services. Under the above perspective about the role of the parameters, we explore the possibilities that guarantee sustainability, and identify associated trade-offs. In particular, we investigate on the actions about the expansion of node cache size K or the growth of network size N, in response to the content expansion. As discussed next in detail, K is the main 'knob' to adjust in counterbalancing M's expansion; investing in memory is straightforward and most likely an inexpensive way to keep the order of C low (ideally bounded) compared to increasing the wireless links' capacity. Complementary to this, expanding network size N can be of help and traded off in place of memory K in some cases, as presented next.

These are applicable both in the centralized paradigm of the network operator controlling K and N, as well as in the adhoc paradigm as the rules of cooperation of the users to preserve the sustainability of the adhoc network.

4.4. Classification of the Scaling Laws

Table 2 summarizes the scaling laws of C for each τ (cases $\tau = 1$ and $\tau = 3/2$ are omitted for simplicity due to similarities up to a slow scaling logarithmic factor with $\tau < 1$ and $\tau > 3/2$, respectively). The laws are grouped in two regimes that lead to networks requiring high and low link capacity, respectively. This grouping will facilitate us to identify interesting trade-offs and draw conclusions about the sustainability of the system.

4.4.1. 'High' Link Capacity Regime

The regime of fast scaling link rate is characterized next as unsustainable, because the increases of network node cache K and/or node count N against content volume M are not sufficient to keep C low. First, note that when the replication storage capacity beyond the storage of the primary copy is limited, i.e., $KN - M = O(1)$, C scales as fast as $\sqrt{N}$, the law of [5]. Observe, moreover, that for $\tau > 1$, $\sqrt{N}$ is scaled down by the capacity $KN - M$ available for replication beyond the primary copy at some power equal to$\sqrt{KN-M}$ or $(KN-M)^{\tau-1}$ for $\tau > 3/2$ or $1 < \tau < 3/2$, respectively; this is precisely this quantifies the gain from adding extra capacity in the network beyond the one required to store a primary copy per file.

On the other hand, for $\tau < 1$, the similarity with the random communicating pairs of [5] is apparent. In the special case of $K = 1$ and $KN = M$, i.e., no capacity left for replication, such setup creates a matching of each node to K uniquely stored files. Thus, there exist $N = M/K$ random communicating pairs of $\Theta\left(\sqrt{N}\right) = \Theta\left(\sqrt{M/K}\right)$ hops on average, creating a likewise link rate C.

4.4.2. 'Low' Link Capacity Regime

In this regime, the required link rate C increases less fast with content volume M; in some cases, it does not increase at all. Hence, this class can be thought as the one for which caching can make a difference in the sustainability of wireless networks. In particular, depending on the order of parameters K, N and M, the required link rate C can be quite low, or—to compare against [5]—increase at an order lower than $\sqrt{N}$ (but it is still open to interpretation whether it is sustainable).

The most interesting regime to explore for perfect sustainability is the one of $C = O(1)$. As the formulas show, to keep C bounded, the hardest case is on $\tau < 1$: node cache capacity K should scale as fast as content volume M. In the intermediate case of $1 < \tau < 3/2$, node capacity K has to scale with M, but slower, at a sublinear power.

The case of $\tau > 3/2$ is quite interesting, as $C = O(1)$ holds true unless content volume M is quite high against $KN^{\frac{3}{2\tau}}$, a quantity of lower order than total capacity KN given that $\tau > 3/2$. It should be thus easier to satisfy the condition. What is more, expanding network nodes N helps in keeping link rate bounded, unlike $\tau < 3/2$ where node cache size K should be expanded for $C = O(1)$. This relates to the extreme disparity in the content of low vs. high popularity. From the perspective of an existing node n, with the addition of more nodes in the network, the non-popular singly replicated content (i.e., of high index m) remains singly replicated, and some of it moves to the new nodes, freeing up cache place in the neighborhood of n; this space is used to replicate highly popular (low index) content more densely. Overall, C decreases, which helps sustainability.

Last, it should be pointed out that τ appears in all expressions of the conditions (in the multiplicative constant or in the exponent, too) in a way such that an increase in its value helps satisfy the condition of the 'Low' Link Capacity Regime.

4.5. Conclusions on Network Sustainability and Trade-offs

Using the above analysis, we now focus on the central question set forth, that is how to guarantee sustainability in an evolving network by means of caching. The derived laws show diverse link rates C diverging as fast as the network diameter $\sqrt{N}$—the Gupta-Kumar law, or remain bounded $O(1)$. Although increases in the raw link capacity may be possible, the cost and effort would likely exceed the alternative of upgrading the memory/storage of the nodes. Under this perspective, the network (be it centralized or community managed) has to control the expansion of parameters K, N and M relative to each other with the goal of keeping the link rates C at a low order (much lower than $\sqrt{N}$), if not bounded.

The 'High Link Capacity' regime correspond to a *cache-deficient* system: C scales fast, as content volume M is close to the total storage capacity KN; incremental cache upgrades have negligible effect on the order of C. Only heavy investment in cache capacity to drive the network to the 'Low Link Capacity' regime would make sense.

The 'Low Link Capacity' regime, on the other hand, correspond to a *cache-efficient* system.However, different values of τ result in much different cases: the smaller the τ, the higher node cache capacity K should increase with content volume M:

- $\tau < 1$: The network should expand node cache capacity K proportionally to content volume M to keep C bounded.

  The network expansion on N is irrelevant, except if the network can tolerate an increasing C (e.g. thanks to incremental increases in link capacity due to advances in the Physical/MAC Layers as the network expands), opting to expand K at a lower rate than M. In such a case, N would just need to fill-in the gap between the two, so as to maintain the condition on the total network storage KN vs. M.

- $1 < \tau < 3/2$: In this case, both expansion in node caches and network nodes can help keep C low against content expansion M. In the scaling rule, the $1/2$ exponent of K is higher than the $3/2 - \tau$ exponent of M (unlike the $\tau < 1$ case); therefore, K can increase at a lower rate than M, with N filling in the gap in view of the condition on the total network storage KN vs. M, as discussed before. Thus, for this case of τ, the network operator can take advantage of the network node expansion and increase node cache size at a slower rate than content volume M.

- $\tau > 3/2$: it is quite easy to keep C bounded (or even decreasing as seen in Figure A) with increases in either the node cache capacity K or the network size N. As suggested from the functional form of the conditions on K and N, the network designer has almost complete freedom to trade cache capacity increases for nodes. However, it should be stressed that such a τ has been observed sparingly.

To link with the measurements on τ reported in the literature, the adverse case with respect to the sustainability of $\tau < 1$ is the more common in the Internet traffic. Values greater than 1, and greater than $3/2$ in particular, have been observed in specialized and particular loads such as mobile traffic [32], UGC [33], or busy servers [31]. It seems, therefore, quite unlikely that a real network would fall in the most favorable case of $\tau > {}^3\!/_2$, but not impossible for the intermediate case of $1 < \tau < {}^3\!/_2$.

In conclusion, increasing the cache size of the nodes has a positive effect in network sustainability as intuitively expected. The same applies to increasing node count, a not as obvious outcome. Quantitatively, the outcome of the two is different and heavily dependent on τ and the relative scaling of the three parameters K, N and M. To ensure network sustainability, the operator/community can exploit the trend of increasing memory in modern devices and strike the desired balance between M against K and N w.r.t. the associated laws.

5. Conclusions & Future Work

In this work, we investigated on the effect of caching in the asymptotic capacity of wireless networks, extending and completing [6, 7] with the full set of scaling laws and a thorough study on the issue of sustainability. Specifically, we conclude that cache expansion is the only means to keep the network sustainable when the Zipf parameter is less than 1. For higher values, sustainability is easier and attainable through a combination of node cache size K expansion and network node expansion N according to the respective trade-offs on each τ. For $\tau > {}^3\!/_2$ in particular, admittedly a non-common case in measured traffic, sustainability is quite easily attainable.

The underlying network model assumed a perfect rectangular grid topology, identical nodes, and multihop communications. Interesting extensions can, hence, investigate on the effect of deviations from the uniformity in the derived laws. Intuitively, one could expect that small deviations preserve the laws, whereas large deviations raise the order towards higher link capacity C. Random topologies and with node mobility have already been considered in a similar context in [18], identifying how much mobility is allowed so that the asymptotic laws of the static case are preserved. Complementary to these is also the direction of users with dissimilar content preferences.

On the subject of the practical implications of this theoretical study, our work makes a step forward on the problem of sustainability of wireless networks, as first set by Gupta and Kumar [5], when networks are enhanced with caching. A similar forward step could research on wireless communications under the paradigm of other advanced techniques, e.g. combining caching with cooperative transmissions in the spirit of [12, 19] and quantifying the gain w.r.t. the scaling laws and network sustainability.

Another interesting extension in this line of research can consider the effect of time-varying content and time-varying popularity $p_m(t)$. On the static content popularity, it suffices to consider static cache allocation and delivery routes, and turns negligible the initial network load overhead to fill-in the caches with content, as this can be amortized over an arbitrarily long network operation. However, when content popularity changes with time, there is a continual load related to updating cache contents, and a transient allocation in the cache contents. It is, therefore, quite important to quantify on how fast popularity can change before new scaling laws about the total load emerge in comparison to the static case.

6. Acknowledgements

The research was supported by Samsung Inc. through the GRO program NETSTAT.

Appendix A. Systemization of the Asymptotic Laws

In this Appendix, we systematize the results of Section 3 in a form practical to understand the scaling of the various quantities and, thus, the system behavior. The laws about the required link capacity C along with the indices l and r are listed in Figure A in a tabular form. In particular, for each case of τ, the Figure depicts two tables, one about C and one about l, r and $\mathcal{M}_{\downarrow}$. In each of them, the various scaling regimes are organized in columns.

A main obstacle to understand the behavior of the system and the association between the laws of C vs. the ones of indices l and r arises from the no one-to-one correspondence between their scaling regimes in the respective Theorems and Lemmas. To overcome this, we subdivide them into finer regimes (shown in shaded columns and marked with gray arrows) where needed (then, matchings are illustrated with gray arrows between the tables).

To facilitate the understanding of the order of the listed laws, we make use of strength bar, e.g. ▂▃▅▆█. Such bars are shown for each quantity C, indices l, r and the conditions on ${}^K\!/_M$ and ${}^{KN}\!/_M$ (the individual node and total network cache capacity normalized over the content volume); naturally, each strength bar is calibrated to the associated quantity, i.e., ▂▯▯▯▯ corresponds to the lowest order possible of $C = o(1)$, $l \to 1$ and $r \to 1$, while ▂▃▅▆█ corresponds to $C = O\left(\sqrt{N}\right)$, $l \sim K$ and $r \to M + 1$, across all cases of τ. As many of the laws can span many different orders, this 'order range' is depicted through the use of gray bars: e.g., as $C = \Theta\left(\sqrt{{}^M\!/_K}\right)$ for $\tau = 1$ can be as low as to $o(1)$ and as high as $O\left(\sqrt{N}\right)$, this is marked with ▂▃▓▓▓.

Regarding the scaling of the link rate C in particular, along with the main law, we present equivalent expressions and bounds (upper and lower) that permit to switch from one subset of the three size parameters to another (e.g. from K and M to K and N for $\tau < 1$). This comes handy in changing the perspective from the content volume M to, for example, the number of nodes N and compare against the Gupta-Kumar $\Theta\left(\sqrt{N}\right)$ law. These follow easily from the main law and the associated scaling regime conditions are thus given without proof.

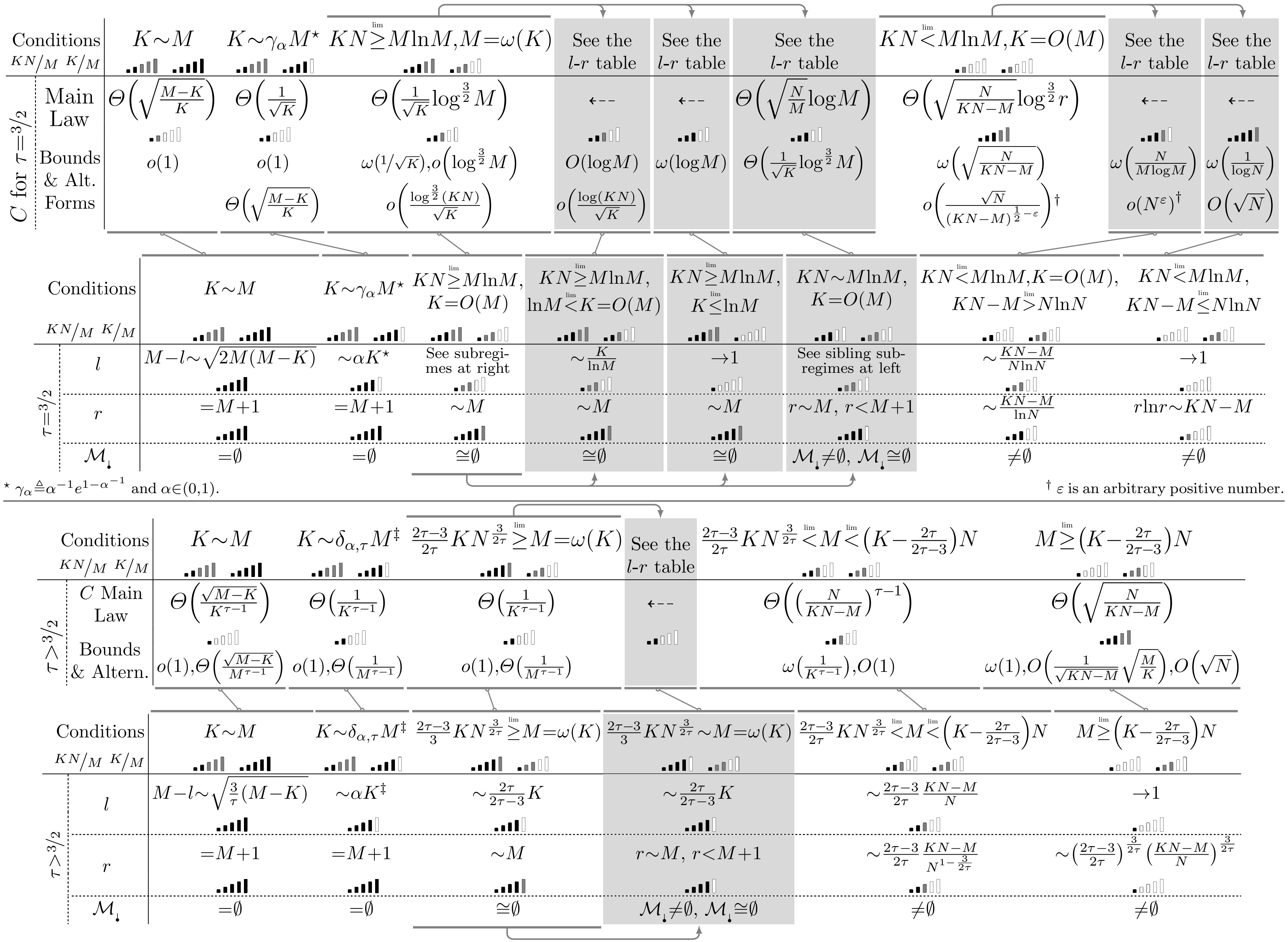
		$K\sim M$	$K\sim\beta_{\alpha,\tau}M^*$	$\frac{3-2\tau}{3}KN\overset{\lim}{\geq}M=\omega(K)$	See the l-r table	See the l-r table	See the l-r table	$\frac{3-2\tau}{3}KN\overset{\lim}{<}M$	See the l-r table	See the l-r table
Conditions $^{KN}/_{M}$ $^{K}/_{M}$		■■▨▨▨ ■■■■■	■■▨▨▨ ■■■■□	■■■▨▨ ■▨▨□□				■▨□□□ ■▨▨□□		
C for $\tau<1$	Main Law	$\Theta\left(\sqrt{\frac{M-K}{K}}\right)$	$\Theta(1)$	$\Theta\left(\sqrt{\frac{M}{K}}\right)$	←--	←--	$\Theta\left(\sqrt{N}\right)$	$\Theta\left(\sqrt{N}\right)$	←--	←--
		■▨□□□	■■■□□	■■■■▨	■■■■□	■■■■▨	■■■■■	■■■■■	■■■■■	■■■■■
	Bounds & Alt. Forms	$o(1)$	$\Theta\left(\sqrt{\frac{M}{K}}\right)$ $\Theta\left(\sqrt{\frac{M-K}{K}}\right)$	$\Theta\left(\sqrt{\frac{M-K}{K}}\right),\Omega\left(K^{\frac{2\tau}{3-2\tau}}\right)$ $\omega(1),O\left(\sqrt{M}\right),O\left(\sqrt{N}\right)$	$O(M^{\frac{\tau}{6}})$ $O\left((KN)^{\frac{\tau}{6}}\right)$	$\omega(M^{\frac{\tau}{6}})$ $\omega(K^{\frac{\tau}{6}})$	$\Theta\left(\sqrt{\frac{M}{K}}\right)$ $\Theta\left(\sqrt{\frac{M-K}{K}}\right)$	$\Theta\left(\sqrt{\frac{M}{K}}\right)$ $\Theta\left(\sqrt{\frac{M-K}{K}}\right)$		
C for $\tau=1$	Main Law	$\Theta\left(\sqrt{\frac{M-K}{K}}\right)$	$\Theta\left(\frac{1}{\log K}\right)$	$\Theta\left(\frac{1}{\log K}\sqrt{\frac{M}{K}}\right)$	←--	←--	$\Theta\left(\frac{\sqrt{N}}{\log K}\right)$	$\Theta\left(\frac{\sqrt{N}}{\log M}\right)$	←--	←--
		■▨□□□	■■■□□	■■■■▨	■■■■□	■■■■▨	■■■■■	■■■■■	■■■■■	■■■■■
	Bounds & Alt. Forms	$o(1)$	$o(1)$ $\Theta\left(\frac{1}{\log K}\sqrt{\frac{M}{K}}\right)$	$\omega\left(\frac{1}{\log K}\right)$ $O\left(\frac{\sqrt{N}}{\log K}\right)$	$O\left(\frac{\sqrt[6]{M}}{\log K}\right)$ $O\left(\frac{\sqrt[6]{KN}}{\log K}\right)$	$\omega\left(\frac{\sqrt[6]{M}}{\log K}\right)$ $\omega\left(\frac{\sqrt[6]{KN}}{\log K}\right)$	$\Theta\left(\frac{1}{\log K}\sqrt{\frac{M}{K}}\right)$	$\Theta\left(\frac{1}{\log M}\sqrt{\frac{M}{K}}\right)$		
C for $1<\tau<{}^3/_2$	Main Law	$\Theta\left(\frac{\sqrt{M-K}}{K^{\tau-1}}\right)$	$\Theta\left(\frac{1}{K^{\tau-1}}\right)$	$\Theta\left(\frac{M^{\frac{3}{2}-\tau}}{\sqrt{K}}\right)$	←--	←--	$\Theta\left(\frac{\sqrt{N}}{M^{\tau-1}}\right)$	$\Theta\left(\frac{\sqrt{N}}{(KN-M)^{\tau-1}}\right)$	←--	←--
		■□□□□	■■□□□	■■■▨□	■■■▨□	■■■▨□	■■■■□	■■■■▨	■■■■□	■■■■■
	Bounds & Alt. Forms	$o(1)$ $\Theta\left(\frac{\sqrt{M-K}}{M^{\tau-1}}\right)$	$o(1),\Theta\left(\frac{\sqrt{M-K}}{K^{\tau-1}}\right)$ $\Theta\left(\frac{1}{M^{\tau-1}}\right),\Theta\left(\frac{M^{\frac{3}{2}-\tau}}{\sqrt{K}}\right)$	$\omega(1),\omega\left(\frac{1}{M^{\tau-1}}\right)$ $\omega\left(\frac{1}{K^{\tau-1}}\right),O\left(\frac{\sqrt{N}}{M^{\tau-1}}\right)$	$O\left(M^{1-\frac{2\tau}{3}}\right)$ $O\left((KN)^{1-\frac{2\tau}{3}}\right)$	$\omega\left(M^{1-\frac{2\tau}{3}}\right)$ $\omega(K)$	$\Theta\left(\frac{M^{\frac{3}{2}-\tau}}{\sqrt{K}}\right)$	$\omega\left(\frac{M^{\frac{3}{2}-\tau}}{\sqrt{K}}\right),O\left(\sqrt{\frac{M}{K}}\right)$ $\omega\left(\frac{\sqrt{N}}{M^{\tau-1}}\right),O\left(\sqrt{N}\right)$	$O\left(\left(\frac{M}{K}\right)^{\frac{3-2\tau}{2\tau}}\right)$ $O\left(N^{\frac{3-2\tau}{2\tau}}\right)$	$\omega\left(\left(\frac{M}{K}\right)^{\frac{3-2\tau}{2\tau}}\right)$ $\omega\left(N^{\frac{3-2\tau}{2\tau}}\right)$

		$K\sim M$	$K\sim\beta_{\alpha,\tau}M^*$	$\frac{3-2\tau}{3}KN\overset{\lim}{\geq}M$, $K=o(M)$	$\frac{3-2\tau}{3}KN\overset{\lim}{\geq}M$, $K\overset{\lim}{\geq}M^{1-\frac{2\tau}{3}}$, $K=o(M)$	$\frac{3-2\tau}{3}KN\overset{\lim}{\geq}M$, $K\overset{\lim}{<}M^{1-\frac{2\tau}{3}}$	$\frac{3-2\tau}{3}KN\sim M$	$\frac{3-2\tau}{3}KN\overset{\lim}{<}M$, $KN-M\overset{\lim}{\geq}\frac{2\tau}{3-2\tau}N^{\frac{3}{2\tau}}$	$\frac{3-2\tau}{3}KN\overset{\lim}{<}M$, $KN-M\overset{\lim}{\leq}\frac{2\tau}{3-2\tau}N^{\frac{3}{2\tau}}$
Conditions $^{KN}/_{M}$ $^{K}/_{M}$		■■▨▨▨ ■■■■■	■■▨▨▨ ■■■■□	■■■▨▨ ■▨▨□□	■■■▨▨ ■■■□□	■■■▨▨ ■▨□□□	■■■□□ ■▨▨□□	■■□□□ ■▨▨□□	■□□□□ ■▨▨□□
$\tau<{}^3/_2$	l	$M-l\sim\sqrt{\frac{3}{\tau}(M-K)}$	$\sim\alpha K^*$	See subregimes at right	$\sim\left[\frac{3-2\tau}{3}\right]^{\frac{3}{2\tau}}\frac{K^{\frac{3}{2\tau}}}{M^{\frac{3}{2\tau}-1}}$	$\to 1$	See sibling subregimes at left	$\sim\frac{3-2\tau}{3}(KN-M)$	$\to 1$
		■■■■■	■■■■□	■▨▨□□	■■▨□□	■□□□□	■▨▨□□	■■□□□	■□□□□
	r	$=M+1$	$=M+1$	$\sim M$	$\sim M$	$\sim M$	$r\sim M$, $r<M+1$	$\sim\frac{3-2\tau}{2\tau}(KN-M)$	$\sim\frac{3-2\tau}{2\tau}(KN-M)$
		■■■■■	■■■■■	■■■■▨	■■■■▨	■■■■▨	■■■■□	■■▨□□	■▨□□□
	$\mathcal{M}_{\downarrow}$	$=\emptyset$	$=\emptyset$	$\cong\emptyset$	$\cong\emptyset$	$\cong\emptyset$	$\mathcal{M}_{\downarrow}\neq\emptyset$, $\mathcal{M}_{\downarrow}\cong\emptyset$	$\neq\emptyset$	$\neq\emptyset$

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

Figure A.1: The scaling laws for C (upper table) and l, r, $\mathcal{M}_{\downarrow}$ (lower table) for $\tau<{}^3/_2$.

Links between the columns of different tables show matchings between the respective regimes. Arrows between columns of the same table indicate subdivisions to finer scaling regimes (shaded columns). Dashed arrow ←-- refers to the entry of the parent regime at the left.Strength bars ■■▨▨▨ depict the law order range; black and gray bars indicate the lower and upper bound of the associated quantity, respectively. E.g., ■□□□□ and ■■■■■ mark the lowest and highest possible orders across all τ, while ■▨▨▨▨ marks a rule that can range from its lowest to highest order.



C for $\tau=3/2$									
Conditions KN/M K/M	$K\sim M$	$K\sim\gamma_\alpha M^\star$	$KN\overset{\lim}{\geq}M\ln M, M=\omega(K)$	See the l-r table	See the l-r table	See the l-r table	$KN\overset{\lim}{<}M\ln M, K=O(M)$	See the l-r table	See the l-r table
Main Law	$\Theta\left(\sqrt{\frac{M-K}{K}}\right)$	$\Theta\left(\frac{1}{\sqrt{K}}\right)$	$\Theta\left(\frac{1}{\sqrt{K}}\log^{\frac{3}{2}}M\right)$	←--	←--	$\Theta\left(\sqrt{\frac{N}{M}}\log M\right)$	$\Theta\left(\sqrt{\frac{N}{KN-M}}\log^{\frac{3}{2}}r\right)$	←--	←--
Bounds & Alt. Forms	$o(1)$	$o(1)$	$\omega(1/\sqrt{K}), o\left(\log^{\frac{3}{2}}M\right)$	$O(\log M)$	$\omega(\log M)$	$\Theta\left(\frac{1}{\sqrt{K}}\log^{\frac{3}{2}}M\right)$	$\omega\left(\sqrt{\frac{N}{KN-M}}\right)$	$\omega\left(\frac{N}{M\log M}\right)$	$\omega\left(\frac{1}{\log N}\right)$
		$\Theta\left(\sqrt{\frac{M-K}{K}}\right)$	$o\left(\frac{\log^{\frac{3}{2}}(KN)}{\sqrt{K}}\right)$	$o\left(\frac{\log(KN)}{\sqrt{K}}\right)$			$o\left(\frac{\sqrt{N}}{(KN-M)^{\frac{1}{2}-\varepsilon}}\right)^\dagger$	$o(N^\varepsilon)^\dagger$	$O\left(\sqrt{N}\right)$

$\tau=3/2$								
Conditions KN/M K/M	$K\sim M$	$K\sim\gamma_\alpha M^\star$	$KN\overset{\lim}{\geq}M\ln M$, $K=O(M)$	$KN\overset{\lim}{\geq}M\ln M$, $\ln M\overset{\lim}{<}K=O(M)$	$KN\overset{\lim}{\geq}M\ln M$, $K\overset{\lim}{\leq}\ln M$	$KN\sim M\ln M$, $K=O(M)$	$KN\overset{\lim}{<}M\ln M, K=O(M)$, $KN-M\overset{\lim}{>}N\ln N$	$KN\overset{\lim}{<}M\ln M$, $KN-M\overset{\lim}{\leq}N\ln N$
l	$M-l\sim\sqrt{2M(M-K)}$	$\sim\alpha K^\star$	See subregimes at right	$\sim\frac{K}{\ln M}$	$\to 1$	See sibling subregimes at left	$\sim\frac{KN-M}{N\ln N}$	$\to 1$
r	$=M+1$	$=M+1$	$\sim M$	$\sim M$	$\sim M$	$r\sim M,\ r<M+1$	$\sim\frac{KN-M}{\ln N}$	$r\ln r\sim KN-M$
$\mathcal{M}_\downarrow$	$=\emptyset$	$=\emptyset$	$\cong\emptyset$	$\cong\emptyset$	$\cong\emptyset$	$\mathcal{M}_\downarrow\neq\emptyset,\ \mathcal{M}_\downarrow\cong\emptyset$	$\neq\emptyset$	$\neq\emptyset$

* $\gamma_\alpha\triangleq\alpha^{-1}e^{1-\alpha^{-1}}$ and $\alpha\in(0,1)$.

† ε is an arbitrary positive number.

$\tau>3/2$						
Conditions KN/M K/M	$K\sim M$	$K\sim\delta_{\alpha,\tau}M^\ddagger$	$\frac{2\tau-3}{2\tau}KN^{\frac{3}{2\tau}}\overset{\lim}{\geq}M=\omega(K)$	See the l-r table	$\frac{2\tau-3}{2\tau}KN^{\frac{3}{2\tau}}\overset{\lim}{<}M\overset{\lim}{<}\left(K-\frac{2\tau}{2\tau-3}\right)N$	$M\overset{\lim}{\geq}\left(K-\frac{2\tau}{2\tau-3}\right)N$
C Main Law	$\Theta\left(\frac{\sqrt{M-K}}{K^{\tau-1}}\right)$	$\Theta\left(\frac{1}{K^{\tau-1}}\right)$	$\Theta\left(\frac{1}{K^{\tau-1}}\right)$	←--	$\Theta\left(\left(\frac{N}{KN-M}\right)^{\tau-1}\right)$	$\Theta\left(\sqrt{\frac{N}{KN-M}}\right)$
Bounds & Altern.	$o(1),\Theta\left(\frac{\sqrt{M-K}}{M^{\tau-1}}\right)$	$o(1),\Theta\left(\frac{1}{M^{\tau-1}}\right)$	$o(1),\Theta\left(\frac{1}{M^{\tau-1}}\right)$		$\omega\left(\frac{1}{K^{\tau-1}}\right),O(1)$	$\omega(1),O\left(\frac{1}{\sqrt{KN-M}}\sqrt{\frac{M}{K}}\right),O\left(\sqrt{N}\right)$

$\tau>3/2$						
Conditions KN/M K/M	$K\sim M$	$K\sim\delta_{\alpha,\tau}M^\ddagger$	$\frac{2\tau-3}{3}KN^{\frac{3}{2\tau}}\overset{\lim}{\geq}M=\omega(K)$	$\frac{2\tau-3}{3}KN^{\frac{3}{2\tau}}\sim M=\omega(K)$	$\frac{2\tau-3}{2\tau}KN^{\frac{3}{2\tau}}\overset{\lim}{<}M\overset{\lim}{<}\left(K-\frac{2\tau}{2\tau-3}\right)N$	$M\overset{\lim}{\geq}\left(K-\frac{2\tau}{2\tau-3}\right)N$
l	$M-l\sim\sqrt{\frac{3}{\tau}(M-K)}$	$\sim\alpha K^\ddagger$	$\sim\frac{2\tau}{2\tau-3}K$	$\sim\frac{2\tau}{2\tau-3}K$	$\sim\frac{2\tau-3}{2\tau}\frac{KN-M}{N}$	$\to 1$
r	$=M+1$	$=M+1$	$\sim M$	$r\sim M,\ r<M+1$	$\sim\frac{2\tau-3}{2\tau}\frac{KN-M}{N^{1-\frac{3}{2\tau}}}$	$\sim\left(\frac{2\tau-3}{2\tau}\right)^{\frac{3}{2\tau}}\left(\frac{KN-M}{N}\right)^{\frac{3}{2\tau}}$
$\mathcal{M}_\downarrow$	$=\emptyset$	$=\emptyset$	$\cong\emptyset$	$\mathcal{M}_\downarrow\neq\emptyset,\ \mathcal{M}_\downarrow\cong\emptyset$	$\neq\emptyset$	$\neq\emptyset$

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

Figure A.2: The scaling laws (cont'd) for $\tau=3/2$ (upper half) and $\tau>3/2$ (lower half).

Appendix A.1. The Role of τ on System Behavior

The main scaling regimes of the tables of Figure A listed in columns from left to right regard the cases of (i) $K \sim M$, (ii) $K = \Theta(M)$, (iii) $\mathcal{M}_{\downarrow} \cong \emptyset$ and (iv) $\mathcal{M}_{\downarrow} \ncong \emptyset$. The first two relate the node cache capacity with the content volume, while the later two pertain to when the number of singly replicated files is negligible or not. In the latter two regimes, for all cases of τ, there exist subcases regarding the law about index l and/or r.

As already discussed in Section 4.2, a lower τ leads to lower order required link rate C; this is manifested in the $K = \Theta(M)$ cases, which is $C = \Theta(1)$ for $\tau < 1$, but $o(1)$ for $\tau \geq 1$, and of even diminishing order as τ increases.

On the subject of the singly replicated files, observe that it becomes increasingly harder to keep non-negligible the cardinality of set $\mathcal{M}_{\downarrow}$ as τ increases. This is intuitively expected; the larger τ implies greater popularity disparity among files. In turn, this means that increasing τ, the low index m files of high popularity become even more popular, while the high index m files' low popularity decreases further, hence, the replication density d_m of the latter is furthered pushed down towards a single copy, expanding set $\mathcal{M}_{\downarrow}$. Regarding indices l and r, the former tends to increase its order (expanding the set $\mathcal{M}_{\uparrow}$ of files replicated at all node), while the latter tends to decrease its order in accordance to set $\mathcal{M}_{\downarrow}$. However, as already discussed, despite the cardinality of $\mathcal{M}_{\downarrow}$ increasing in τ, the order of C falls, which makes possible to keep link rate bounded, $C = O(1)$ on $\tau > 3/2$ even when $\mathcal{M}_{\downarrow} \ncong \emptyset$. Reversely, for $\tau < 1$, link rate C attains the Gupta-Kumar $\sqrt{N}$ law even if $\mathcal{M}_{\downarrow}$ is almost empty, provided that it is not completely empty, i.e. $\mathcal{M}_{\downarrow} \cong \emptyset$, $\mathcal{M}_{\downarrow} \neq \emptyset$.

Appendix B. Proofs

In the following, the underbracket notation marks the significant quantities that carry on to the next step, e.g. $\underbrace{x}_{}+y$ signifies that only x carries on (as e.g., $y = o(x)$).

Proof of Lemma 1. To compute l, we use (9); to find the conditions for $\mathcal{M}_{\downarrow} \approx \emptyset \Leftrightarrow M - r = o(M) \Leftrightarrow r \sim M$, we use (10) when $l = o(K)$, or (13) when $l = \Theta(K)$. If $M_{\downarrow} = \emptyset$, then (10) must be true for $r = M + 1$.
Case $\tau < 3/2$: Using (6) and $r \sim M$ in (9) and (10), it is

$$K - l + 1 - \frac{M - r + 1}{N} \sim l^{\frac{2\tau}{3}} \frac{M^{\frac{3-2\tau}{3}} - (l-1)^{\frac{3-2\tau}{3}}}{1 - \frac{2\tau}{3}}. \qquad (1)$$

$$(K - l + 1)N - (M - r) - 1 \overset{\lim}{\geq} M^{\frac{2\tau}{3}} \frac{M^{\frac{3-2\tau}{3}} - (l-1)^{\frac{3-2\tau}{3}}}{1 - \frac{2\tau}{3}}. \qquad (2)$$

Assuming $l \sim \alpha K$, with $\alpha \in (0, 1)$, results from (1) to

$$\underbrace{(1-\alpha)K}_{} + 1 - \frac{o(M)}{N} \sim \underbrace{(\alpha K)^{\frac{2\tau}{3}}}_{} \frac{\underbrace{M^{\frac{3-2\tau}{3}}}_{} - \underbrace{(\alpha K)^{\frac{3-2\tau}{3}}}_{}}{1 - \frac{2\tau}{3}} \Rightarrow$$

$$M \sim \alpha^{-\frac{2\tau}{3-2\tau}} \left(\frac{3 - 2\tau(1-\alpha)}{3} \right)^{\frac{3}{3-2\tau}} K,$$

which proves the third case for l.

Assuming $l \sim K$, (1) leads, using $M - r = o(M)$, to

$$o(K) + 1 - \frac{o(M)}{N} \sim K^{\frac{2\tau}{3}} \frac{M^{\frac{3-2\tau}{3}} - K^{\frac{3-2\tau}{3}}}{1 - \frac{2\tau}{3}} \Rightarrow$$

$$\underbrace{M^{\frac{3-2\tau}{3}}}_{} \sim \underbrace{K^{\frac{3-2\tau}{3}}}_{} + o\left(K^{\frac{3-2\tau}{3}}\right) - o\left(\frac{M}{K^{\frac{2\tau}{3}} N}\right) \Rightarrow M \sim K,$$

where we use $\frac{M}{K^{\frac{2\tau}{3}} N} = K^{\frac{3-2\tau}{3}} \frac{M}{KN} \overset{(1)}{<} K^{\frac{3-2\tau}{3}}$ in the last step. Note that the rule of the third case applies for $\alpha = 1$.

Examining $\mathcal{M}_{\downarrow}$ in the two above cases of $l = \Theta(K) = \Theta(M)$, and assuming $r \leq M$ (i.e. $\mathcal{M}_{\downarrow}$ not strictly empty), we can use (13) to get $r = \Theta\left(KN^{\frac{3}{2\tau}}\right)$, which is $\omega(M)$, provided that $N = \omega(1)$. This is a contradiction as $r \leq M + 1$. Thus, $r = M + 1$, and $\mathcal{M}_{\downarrow} = \emptyset$. This completes the third case of the conditions for almost empty $\mathcal{M}_{\downarrow}$.

To complete the fourth cases of l with the estimate of $M - l$, we do a Taylor expansion to three terms of function $f(x) = (1-x)^{\beta}$. For $x \cong 0$, it is $f(x) \cong 1 - \beta x + \beta \frac{\beta - 1}{2} x^2$. Using the fact that $r = M + 1$, we apply (1) again:

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

$$\left(1 - \frac{2\tau}{3}\right)\left(\frac{K}{M} - \frac{l}{M} + \frac{1}{M}\right) \cong \left[\frac{l}{M}\right]^{\frac{2\tau}{3}} - \frac{l}{M} \Rightarrow$$

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

Using Taylor's approximation of $f(x)$ for $\beta = \frac{2\tau}{3}$ and $x = \frac{M-l}{M}$, it is $\left[\frac{l}{M}\right]^{\frac{2\tau}{3}} \cong 1 - \frac{2\tau}{3} \frac{M-l}{M} + \frac{2\tau}{3} \frac{2\tau - 3}{6} \left[\frac{M-l}{M}\right]^2$. Substituting,

$$1 - \frac{2\tau}{3} - \frac{2\tau}{3} \frac{3 - 2\tau}{6} \left[\frac{M-l}{M}\right]^2 \cong \left(1 - \frac{2\tau}{3}\right) \frac{K+1}{M} \Rightarrow$$

$$M - l \cong M \sqrt{\frac{3}{\tau} \frac{M - K - 1}{M}},$$

which is valid when $M - K = \omega(1)$. This completes the first case of l.

Turning our attention to the other two cases of l, i.e., when $K = o(M)$, it must be $l = o(K)$ since $l = \Theta(K)$ leads to the above case. Evidently $l = o(M)$ and (1) becomes

$$\underbrace{K}_{} - \frac{o(M)}{N} \sim \underbrace{l^{\frac{2\tau}{3}}}_{} \frac{\underbrace{M^{1 - \frac{2\tau}{3}}}_{}}{1 - \frac{2\tau}{3}} \Rightarrow l \sim \left(1 - \frac{2\tau}{3}\right)^{\frac{3}{2\tau}} \frac{K^{\frac{3}{2\tau}}}{M^{\frac{3}{2\tau} - 1}}.$$

This is correct (second case of l) as long as $l \overset{\lim}{>} 1$, i.e. if $\left(1 - \frac{2\tau}{3}\right)^{\frac{3}{2\tau}} K^{\frac{3}{2\tau}} \overset{\lim}{\geq} M^{\frac{3}{2\tau} - 1}$. Otherwise, (9) was not applicable, thus $l \to 1$.

Last, for the last two cases of $l = o(K)$, (2) assures $\mathcal{M}_{\downarrow} \approx \emptyset$ when $M \overset{\lim}{\leq} \left(1 - \frac{2\tau}{3}\right) KN$; if the inequality is strict, $\mathcal{M}_{\downarrow} = \emptyset$. However, as we assumed $l = o(K)$, we should add the constraint of $K = o(M)$ or equivalently $M = \omega(K)$. These complete the first two cases of the conditions on almost empty $\mathcal{M}_{\downarrow}$.

Case $\tau = 3/2$: Similarly, (6) and $M - r = o(M)$ in (9)-(10) lead to

$$K - l + 1 - \frac{M - r + 1}{N} \sim l \ln \frac{M}{l-1} \tag{3}$$

$$(K - l + 1)N - (M - r) - 1 \overset{\lim}{\geq} M \ln \frac{M}{l-1} \tag{4}$$

Assuming $l \sim K$, (3) becomes

$$\begin{aligned} \underbrace{o(K)}_{} + 1 - o\left(\frac{M}{N}\right) \sim & \quad \underbrace{K}_{} (\underbrace{\ln M}_{} - \underbrace{\ln K}_{}) \overset{\frac{M}{KN}<1}{\Rightarrow} \\ \ln M \sim & \quad \ln K + o(1) \Rightarrow M \sim K, \end{aligned}$$

Assuming $\mathcal{M}_{\downarrow} \neq \emptyset$ as before, we can use (13) to get $r \sim KN$, which is a contradiction, provided that $N = \omega(1)$. Thus, $\mathcal{M}_{\downarrow} = \emptyset$. This completes the first case of $\mathcal{M}_{\downarrow}$.

For $l \sim \alpha K$, with $\alpha \in (0, 1)$, (3) becomes

$$\underbrace{(1-\alpha)K}_{} + 1 + o(K) - o\left(\frac{M}{N}\right) \sim \underbrace{\alpha K}_{} \underbrace{\ln \frac{M}{\alpha K}}_{} \Rightarrow M \sim \alpha K e^{\frac{1-\alpha}{\alpha}}.$$

From (1), it is $\frac{M}{KN} \to 0$; this completes the third case of l.

To derive the condition of $\mathcal{M}_{\downarrow} \cong \emptyset$, we substitute $l \sim K \sim M$ in (4): it follows that $e^{\frac{\alpha-1}{\alpha}} N \overset{\lim}{\geq} 1$, which is a strict inequality provided $N = \omega(1)$; this concludes the second and third cases of $\mathcal{M}_{\downarrow}$.

Returning to the case of $l \sim K$, we estimate $M - l$ by applying the Taylor's approximation $f(1 + x) \triangleq \ln x \cong x - \frac{1}{2}x^2$. Rewriting (3) with $r = M + 1$,

$$\begin{aligned} K - l + 1 \cong l \ln \frac{M}{l} \cong l \left(\frac{M - l}{l} - \frac{1}{2} \left[\frac{M - l}{l} \right]^2 \right) \Rightarrow \\ 0 \cong l^2 - 2(2M - K - 1)l + M^2 \Rightarrow \\ l \cong 2M - K - 1 - M\sqrt{\left[1 + \frac{M - K - 1}{M}\right]^2 - 1} \Rightarrow \\ l \cong 2M - K - 1 - M\sqrt{2\frac{M - K - 1}{M}}, \end{aligned}$$

where the last step uses $(1+x)^2 \cong 1+2x$ for $x \to 0$. Then,

$$M - l = \underbrace{M\sqrt{2\frac{M-K-1}{M}}}_{} - (M-K-1) \sim M\sqrt{2\frac{M-K-1}{M}},$$

when $M - K = \omega(1)$.

For the other two cases of l and $\mathcal{M}_{\downarrow}$, it is $l = o(K) = o(M)$, and given that $K = O(M)$ from (1),

$$\underbrace{K}_{} - l + 1 - \frac{M - r + 1}{N} \sim \underbrace{l}_{} (\underbrace{\ln M}_{} - \ln l) \Rightarrow l \sim \frac{K}{\ln M}.$$

The above (second case for l) is true if $\ln M \overset{\lim}{<} K = o(M)$, otherwise $l \to 1$ (first case) as (9) is not applicable.

Last, (4) leads to $KN \overset{\lim}{\geq} M \ln M$ to have $\mathcal{M}_{\downarrow} \approx \emptyset$.

Case $\tau > 3/2$: (1) and (2) are applicable in this case, too, and they can be rewritten as

$$K - l + 1 - \frac{M - r + 1}{N} \sim l^{\frac{2\tau}{3}} \frac{\left[\frac{1}{l-1}\right]^{\frac{2\tau}{3}-1} - \left[\frac{1}{M}\right]^{\frac{2\tau}{3}-1}}{\frac{2\tau}{3} - 1}. \tag{5}$$

$$(K - l + 1)N - (M - r) - 1 \overset{\lim}{\geq} M^{\frac{2\tau}{3}} \frac{\left[\frac{1}{l-1}\right]^{\frac{2\tau}{3}-1} - \left[\frac{1}{M}\right]^{\frac{2\tau}{3}-1}}{\frac{2\tau}{3} - 1}. \tag{6}$$

Assuming $l \sim K$, the same derivation as in $\tau < 3/2$ leads to $K \sim M$. Repeating, moreover, the derivation for $l \sim \alpha K$ as in $\tau < 3/2$, we have to take care to ensure that $3 - 2\tau(1 - \alpha) > 0$, or equivalently, $\alpha > \frac{2\tau-3}{2\tau}$, which is the second case of l. Note, that in the previous cases this constraint was automatically satisfied.

Repeating the analysis on $\mathcal{M}_{\downarrow}$ as in the two above cases of $l = \Theta(K) = \Theta(M)$, and assuming $\mathcal{M}_{\downarrow}$ is non-empty, leads from (13) to $r \sim KN^{\frac{3}{2\tau}}$, which is $\omega(M)$ (provided that $N = \omega(1)$). This is a contradiction, thus $\mathcal{M}_{\downarrow} = \emptyset$. This completes the third case of the conditions on $\mathcal{M}_{\downarrow}$.

Note that the range of $\alpha \in \left(\frac{2\tau-3}{2\tau}, 1\right]$ covers all the cases of $K = \Theta(M)$. Hence, the last case to consider is $K = o(M)$: then, $l = O(K) = o(M)$, thus (1) leads to

$$\underbrace{K}_{} - \underbrace{l}_{} - \frac{o(M)}{N} \sim \underbrace{l^{\frac{2\tau}{3}}}_{} \frac{\underbrace{(l-1)^{1-\frac{2\tau}{3}}}_{} - M^{1-\frac{2\tau}{3}}}{\frac{2\tau}{3} - 1} \Rightarrow l \sim \frac{2\tau - 3}{2\tau} K,$$

which is the first case of l. Last, with this l, (6) leads to

$$\begin{aligned} \frac{3}{2\tau} \underbrace{KN}_{} - (M - r) - 1 \overset{\lim}{\geq} \underbrace{M^{\frac{2\tau}{3}}}_{} \frac{\underbrace{\left(\frac{2\tau-3}{2\tau}K\right)^{1-\frac{2\tau}{3}}}_{} - M^{1-\frac{2\tau}{3}}}{\frac{2\tau}{3} - 1} \\ \Rightarrow M \overset{\lim}{\leq} \frac{2\tau-3}{2\tau} KN^{\frac{3}{2\tau}}, \end{aligned}$$

which assures $\mathcal{M}_{\downarrow} \approx \emptyset$; if the inequality is strict, $\mathcal{M}_{\downarrow} = \emptyset$.

Similarly, the derivation of $\tau < 3/2$ applies for $M - l$.

Assuming $l = o(K)$ leads through (9) to $l = (\frac{2\tau}{3} - 1)K$ which is a contradiction, thus it is always $l = \Theta(K)$. □

Proof of Lemma 2. First, assume $KN - M = O(1)$; as $KN - M$ are the slots available for replication beyond the primary copy, $r = \Theta(1)$, and thus $l = \Theta(1)$. If $l > 1$, (13) becomes $r = \Theta(N^{\frac{3}{2\tau}})$ which is a contradiction. Thus $l = 1$. Using (12), we compute r.

Next, we proceed to $\mathcal{M}_{\downarrow} \neq \emptyset$ with $KN - M = \omega(1)$, and identify the following cases:

Case $\tau < 3/2$: Assuming that $l \overset{\lim}{>} 1$, we can invoke (13) to substitute r in (9), and get with the help of (6)

$$\begin{aligned} \underbrace{K}_{} - l + 1 - \frac{\underbrace{M}_{} - \underbrace{lN^{\frac{3}{2\tau}}}_{} + 1}{\underbrace{N}_{}} \sim \frac{3\underbrace{l^{\frac{2\tau}{3}}}_{}}{3-2\tau} \left(\underbrace{l^{1-\frac{2\tau}{3}} N^{\frac{3}{2\tau}-1}}_{} - l^{1-\frac{2\tau}{3}}\right) \Rightarrow \\ l \sim \frac{3 - 2\tau}{2\tau} \frac{KN - M}{N^{\frac{3}{2\tau}}} \end{aligned}$$

Then, invoking back (13), $r \sim \frac{3-2\tau}{2\tau}\,(KN-M)$. Clearly, the condition for these to hold true will be $l \overset{\lim}{>} 1$, or equivalently $KN-M \overset{\lim}{>} \frac{2\tau}{3-2\tau}N^{\frac{3}{2\tau}}$. Otherwise, we conclude $l \to 1$ and calculate the scaling of r from (12):

$$\underbrace{KN}_{}-\underbrace{M}_{}+\underbrace{r}_{}-1 \sim \frac{3}{3-2\tau}\,\underbrace{r^{\frac{2\tau}{3}}}_{}\left(\underbrace{r^{1-\frac{2\tau}{3}}}_{}-1\right) \Rightarrow$$
$$r \sim \frac{3-2\tau}{2\tau}(KN-M).$$

Case $\tau = 3/2$: Assuming $l \overset{\lim}{>} 1$, using (6), (9) leads to

$$\underbrace{K}_{}-l+1-\frac{\underbrace{M}_{}-lN+1}{\underbrace{N}_{}} \sim \underbrace{l}_{}\underbrace{\ln\frac{lN}{l}}_{} \Rightarrow l \sim \frac{KN-M}{N\ln N},$$

and (13) to $r \sim \frac{KN-M}{\ln N}$; for these to be true, it has to be $KN-M \overset{\lim}{>} N\ln N$. Otherwise, $l \to 1$, and (12) results in $r\ln r \sim KN-M$.

Case $\tau > 3/2$: Assuming $l \overset{\lim}{>} 1$, using (6) and (13), (9) becomes

$$\underbrace{K}_{}-\underbrace{l}_{}+1-\frac{\underbrace{M}_{}-lN^{\frac{3}{2\tau}}+1}{\underbrace{N}_{}} \sim \frac{3\,\underbrace{l^{\frac{2\tau}{3}}}_{}}{2\tau-3}\left(\underbrace{l^{1-\frac{2\tau}{3}}}_{}-l^{1-\frac{2\tau}{3}}N^{\frac{3}{2\tau}-1}\right) \Rightarrow$$
$$l \sim \frac{2\tau-3}{2\tau}\left(K-\frac{M}{N}\right)$$

and thus, $r \sim \frac{2\tau-3}{2\tau}\frac{KN-M}{N^{1-\frac{3}{2\tau}}}$, provided that $KN-M \overset{\lim}{>} \frac{2\tau}{2\tau-3}N$. Otherwise, $l \to 1$, and from (12), $\underbrace{KN}_{}-\underbrace{M}_{}+r-1 \sim \frac{3}{2\tau-3}\left[\underbrace{r^{\frac{2\tau}{3}}}_{}-r\right] \Rightarrow r \sim \left[\frac{2\tau-3}{2\tau}(KN-M)\right]^{\frac{3}{2\tau}}$. □

Proof of Lemma 3. Since $d_m \geq \frac{1}{N}$, we have $C \triangleq \sum_{m\in\mathcal{M}}\left[\frac{1}{\sqrt{d_m}}-1\right]p_m < \sum_{m\in\mathcal{M}}p_m\sqrt{N} = \sqrt{N}$. □

Proof of Theorem 4 ($K \overset{\lim}{<} M$). First, we study the case of $K \overset{\lim}{<} M$. From Lemma 1, it is $K-l+1 = \Theta(K)$ except on the case $\alpha = 1$ (treated separately at the end) and $r = \Theta(M)$; these yield $\sqrt{K-l+1-\frac{M-r+1}{N}} = \Theta(\sqrt{K})$. Thus, the scaling of $C_{\multimapbothvert}$ is determined from the ratio $\frac{[H_{2\tau/3}(M)-H_{2\tau/3}(K)]^{\frac{3}{2}}}{H_\tau(M)}$. As for $C_{\downarrow}$, we use that for $r = \Theta(M)$, it is $H_\tau(M)-H_\tau(r-1) = \Theta(M^{-\tau}(M-r)) = O\big(M^{1-\tau}\big)$; combined with (5) yields $C_{\downarrow} = o\Big(\sqrt{N}\frac{M^{1-\tau}}{H_\tau(M)}\Big)$. This is used on the cases $\tau < 3/2$. Next, we examine each case separately and show that it is always $C_{\downarrow} = O(C_{\multimapbothvert})$, which implies that $C_{\multimapbothvert}$ alone determines the asymptotic law.

Case $\tau < 1$: We start with the case of $l \sim \alpha K$ with $\alpha \in (0,1)$:

$$C_{\multimapbothvert} = \Theta\left(\frac{[H_{\frac{2\tau}{3}}(M)-H_{\frac{2\tau}{3}}(l)]^{\frac{3}{2}}}{\sqrt{K}H_\tau(M)}\right) = \Theta\left(\frac{[M^{\frac{3-2\tau}{3}}-l^{\frac{3-2\tau}{3}}]^{\frac{3}{2}}}{\sqrt{K}M^{1-\tau}}\right)$$
$$= \Theta\left(\sqrt{\frac{M}{K}}\left[1-\left(\frac{l}{M}\right)^{\frac{3-2\tau}{3}}\right]^{\frac{3}{2}}\right) = \Theta\left(\sqrt{\frac{M}{K}}\right),$$

since $\lim\frac{l}{M} < \lim\frac{l}{K} = \alpha < 1$.

For the other two cases of Lemma 1 on l, it is $l = o(M)$ and $\frac{[H_{2\tau/3}(M)]^{\frac{3}{2}}}{H_\tau(M)} \to \sqrt{M}$; using (4), it is $C_{\multimapbothvert} = \Theta\Big(\sqrt{\frac{M}{K}}\Big)$.

From Lemma 1, if $\mathcal{M}_{\downarrow} = \emptyset$, then $C_{\downarrow} = 0$. Else, $\frac{3-2\tau}{3}KN \sim M$ and $C_{\downarrow} = o\big(\sqrt{N}\big) = o\Big(\sqrt{\frac{M}{K}}\Big) = o\big(C_{\multimapbothvert}\big)$.

Case $\tau = 1$. The first case of $l \sim \alpha K$ is covered from the above derivation, except $H_\tau(M)$ in the denominator, which changes from $M^{1-\tau}$ to $\log M$. Thus,

$$C_{\multimapbothvert} = \Theta\left(\frac{[H_{\frac{2\tau}{3}}(M)-H_{\frac{2\tau}{3}}(l)]^{\frac{3}{2}}}{\sqrt{K}H_\tau(M)}\right) = \Theta\left(\frac{\sqrt{M}}{\sqrt{K}\log M}\right).$$

If l is given by any of the other two cases, $l = o(M)$ and $\frac{[H_{2\tau/3}(M)]^{\frac{3}{2}}}{H_\tau(M)} \to \frac{\sqrt{M}}{\log M}$. Then, (4) leads to $C_{\multimapbothvert} = \Theta(\frac{\sqrt{M}}{\sqrt{K}\log M})$. As before, if $\mathcal{M}_{\downarrow} = \emptyset$, $C_{\downarrow} = 0$. Otherwise, $\frac{3-2\tau}{3}KN \sim M$ and $C_{\downarrow} = o\Big(\frac{\sqrt{N}}{\log M}\Big) = o\Big(\frac{\sqrt{M}}{\sqrt{K}\log M}\Big) = o\big(C_{\multimapbothvert}\big)$.

Case $1 < \tau < 3/2$: Similarly, if $l \sim \alpha K$ with $\alpha \in (0,1)$, we use the above derivations to find

$$C_{\multimapbothvert} = \Theta\Big([H_{\frac{2\tau}{3}}(M)-H_{2\tau/3}(l)]^{\frac{3}{2}}/\sqrt{K}H_\tau(M)\Big) = \Theta\left(\frac{M^{\frac{3}{2}-\tau}}{\sqrt{K}}\right),$$

If l is given by any of the other two cases, $l = o(M)$ and $\frac{[H_{2\tau/3}(M)]^{3/2}}{H_\tau(M)} \to M^{\frac{3}{2}-\tau}$, which from (4), leads to $C_{\multimapbothvert} = \Theta\left(\frac{M^{\frac{3}{2}-\tau}}{\sqrt{K}}\right)$. Similarly as above, if $C_{\downarrow} \neq 0$ then $\frac{3-2\tau}{3}KN \sim M$ and $C_{\downarrow} = o\big(\sqrt{N}M^{1-\tau}\big) = o\left(\frac{M^{\frac{3}{2}-\tau}}{\sqrt{K}}\right) = o\big(C_{\multimapbothvert}\big)$.

Case $\tau = 3/2$: First, let $l \sim \alpha K$ with $\alpha \in (0,1)$; then,

$$\frac{[H_{\frac{2\tau}{3}}(M)-H_{2\tau/3}(l)]^{3/2}}{\sqrt{K}H_\tau(M)} = \Theta\left(\frac{[\ln M-\ln l]^{\frac{3}{2}}}{\sqrt{K}}\right)$$
$$= \Theta\left(\frac{[\ln M-\ln e^{\frac{\alpha-1}{\alpha}}M]^{\frac{3}{2}}}{\sqrt{K}}\right) = \Theta\left(\frac{1-\alpha}{\alpha\sqrt{K}}\right) = \Theta\left(\frac{1}{\sqrt{K}}\right).$$

On this l, it is $K = \Theta(M)$,thus $\mathcal{M}_{\downarrow} = \emptyset$, and $C_{\downarrow} = 0$.

If l is given from the two other cases of Lemma 1, it is $l = o(M)$, hence, $\frac{[H_{2\tau/3}(M)]^{\frac{3}{2}}}{H_\tau(M)} \to \log^{\frac{3}{2}}M$. Then, from (4), $C_{\multimapbothvert} = \Theta\left(\frac{\log^{\frac{3}{2}}M}{\sqrt{K}}\right)$. As before, if $C_{\downarrow} \neq 0$, $KN \sim M\log M$ and $C_{\downarrow} = o\left(\frac{\sqrt{N}}{\sqrt{M}}\right) = o\left(\frac{\log^{\frac{1}{2}}M}{\sqrt{K}}\right) = o\left(\frac{\log^{\frac{3}{2}}M}{\sqrt{K}}\right) = o(C_{\multimapbothvert})$.

Case $\tau > 3/2$: Here, $H_\tau(M)$ are all constants, while $H_{2\tau/3}(r)-H_{2\tau/3}(l) = \Theta\Big(K^{\frac{2\tau}{3}}\Big)$, as $l = \Theta(K)$ and $r \sim M = \omega(K)$. Substituting to (4) leads to $C_{\multimapbothvert} = \Theta\big(1/K^{\tau-1}\big)$.

Similarly, if $C_{\downarrow} > 0$, then $\frac{2\tau-3}{2\tau}KN^{\frac{3}{2\tau}} \sim M$ and $C_{\downarrow} = o\big(\sqrt{N}M^{1-\tau}\big) = o\Big(\frac{1}{K^{\frac{\tau}{3}}M^{\frac{2\tau}{3}-1}}\Big) \overset{K=o(M)}{=} o\big(\frac{1}{K^{\tau-1}}\big) = o(C_{\multimapbothvert})$. □

Proof of Theorem 4 (case of $K \sim M$). For the case of $K \sim M$, it is $\mathcal{M}_{\downarrow} = \emptyset$ (from Lemma 1), thus $C_{\downarrow} = 0$.

Case $\tau <$ 3/2: Equation (1) can be rewritten as

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

where the approximation comes from the Taylor expansion of function $(x+1)^{\beta}-1 \approx \beta x$ for $x = \frac{M}{l}-1 \to 0$ (as $l \sim K \sim M$), and $\beta = 1-\frac{2\tau}{3}$. Replacing $K-l+1$ in $C_{\updownarrow}$,

$$C_{\updownarrow} = \Theta\left(\frac{\left((M-l)M^{-\frac{2\tau}{3}}\right)^{\frac{3}{2}}}{\sqrt{K-l+1}\ H_{\tau}(M)}\right) = \Theta\left(\frac{(M-l)^{\frac{3}{2}}M^{-\tau}}{\sqrt{M-l}\ M^{1-\tau}}\right)$$
$$= \Theta\left(\frac{M-l}{M}\right) \overset{M-l \overset{\lim}{\propto} M\sqrt{\frac{M-K}{M}}}{=\!=\!=} \Theta\left(\sqrt{\frac{M-K}{M}}\right).$$

Case $\tau =$ 3/2: Using Taylor approximation for the logarithm function $\ln(1+x)$ in (3) around $x=0$,

$$K-l+1 = l\ln\frac{M}{l-1} \cong l\left(\frac{M}{l-1}-1\right) \sim M-l.$$

Replacing then $K-l+1$ in $C_{\updownarrow}$,

$$C_{\updownarrow} = \Theta\left(\frac{((M-l)M^{-1})^{\frac{3}{2}}}{\sqrt{K-l+1}\ H_{3/2}(M)}\right) = \Theta\left(\frac{(M-l)^{\frac{3}{2}}M^{-\frac{3}{2}}}{\sqrt{M-l}M^{-\frac{1}{2}}}\right)$$
$$= \Theta\left(\frac{M-l}{M}\right) \overset{M-l \overset{\lim}{\propto} M\sqrt{\frac{M-K}{M}}}{=\!=\!=} \Theta\left(\sqrt{\frac{M-K}{M}}\right)$$

From Lemma 1, it is $\mathcal{M}_{\downarrow} = \emptyset$, thus $C_{\downarrow} = 0$. Substituting thus $M-l$ in $C_{\updownarrow}$, the result follows.

Case $\tau >$ 3/2: The approximation of $\tau <$ 3/2 and (1) apply, leading to $K-l+1 \cong M-l$. Replacing in $C_{\updownarrow}$,

$$C_{\updownarrow} = \Theta\left(\frac{\left((M-l)M^{-\frac{2\tau}{3}}\right)^{\frac{3}{2}}}{\sqrt{K-l+1}\ H_{\tau}(M)}\right) = \Theta\left(\frac{(M-l)^{\frac{3}{2}}M^{-\tau}}{\sqrt{M-l}}\right)$$
$$= \Theta\left(\frac{M-l}{M^{\tau}}\right) \overset{M-l \sim M\sqrt{\frac{3}{\tau}(M-K)}}{=} \Theta\left(\frac{\sqrt{M-K}}{M^{\tau-1}}\right). \qquad \square$$

Proof of Theorem 5. When $\mathcal{M}_{\downarrow} \neq \emptyset$, for all τ, it is $l = o(M)$; indeed, if we assume that $l \not\to 1$, it is $l = rN^{\frac{3}{2\tau}} < MN^{\frac{3}{2\tau}} = o(M)$; if we assume that $l \to 1$, it is similarly $l = o(M)$. Moreover, for all τ, it is $M-r = \Theta(M)$, thus $\sqrt{K-l+1-\frac{M-r+1}{N}} = \Theta\left(\sqrt{K+1-\frac{M}{N}}\right)$.

Case $\tau < 1$: As $M-r = \Theta(M)$, it is $r \overset{\lim}{<} M$. Then, from (6), $C_{\downarrow} = \sqrt{N}\frac{H_{\tau}(M)}{H_{\tau}(M)} = \Theta(\sqrt{N})$. In total, $C = \Theta(\sqrt{N})$.

Case $\tau = 1$: Lemma 1 implies that for $\mathcal{M}_{\downarrow} \neq \emptyset$, it should be $M \overset{\lim}{>} \frac{1}{3}KN$. Also, from Lemma 2, it is $r \sim \frac{1}{2}(KN-M)$. Using $l = o(r)$, it is $C_{\updownarrow} = \Theta\left(\frac{\sqrt{r}}{\sqrt{K-\frac{M}{N}}\log M}\right) = \Theta\left(\frac{\sqrt{N}}{\log M}\right)$.

Next we discern two cases for the relation between r and M that give different laws for $C_{\downarrow}$:

- if $M \sim KN$, Lemma 2 yields $r = o(M)$. Thus, $C_{\downarrow} = \Theta\left(\sqrt{N}\frac{H_{\tau}(M)}{H_{\tau}(M)}\right) = \Theta\left(\sqrt{N}\right)$. In total, $C = \Theta(\sqrt{N})$.
- if $M \overset{\lim}{<} KN$, then $r \sim \beta M$ with $0 < \beta \leq 1/3$, and thus, $C_{\downarrow} = \Theta\left(\sqrt{N}\frac{\log\frac{M}{r}}{\log M}\right) = \Theta\left(\frac{\sqrt{N}}{\log M}\right)$. In total, $C = \Theta\left(\frac{\sqrt{N}}{\log M}\right)$.

Since it is $\frac{1}{3}KN \overset{\lim}{<} M$ for $\mathcal{M}_{\downarrow} \not\approx \emptyset$, no other case exists.

Case $1 < \tau <$ 3/2: As before, $r \sim \frac{3-2\tau}{2\tau}(KN-M)$. Then, $C_{\updownarrow} = \Theta\left(\frac{r^{\frac{3}{2}-\tau}}{\sqrt{K-\frac{M}{N}}}\right) = \Theta\left(\frac{\sqrt{N}}{(KN-M)^{\tau-1}}\right)$.

Moreover, Lemma 2 implies that $r \overset{\lim}{<} M$. Thus, $C_{\downarrow} = \Theta\left(\sqrt{N}[H_{\tau}(M)-H_{\tau}(r)]\right) = \Theta\left(\frac{\sqrt{N}}{r^{\tau-1}}\right) = \Theta\left(\frac{\sqrt{N}}{(KN-M)^{\tau-1}}\right)$. In total, $C = \Theta\left(\frac{\sqrt{N}}{(KN-M)^{\tau-1}}\right)$.

Case $\tau =$ 3/2: As $l = o(r)$, it is $C_{\updownarrow} = \Theta\left(\frac{\log^{\frac{3}{2}} r}{\sqrt{K-\frac{M}{N}}}\right)$. Moreover, $C_{\downarrow} = \Theta\left(\sqrt{N}\left[\frac{1}{\sqrt{r}}-\frac{1}{\sqrt{M}}\right]\right) = O\left(\sqrt{\frac{N}{r}}\right)$. Now, we examine the asymptotic law of r from Lemma 2:

- if $KN-M \overset{\lim}{\leq} N\ln N$, then $r\ln r \sim KN-M$. Then, $C_{\downarrow} = O\left(\sqrt{\frac{N}{r}}\right) = O\left(\sqrt{\frac{N\log r}{KN-M}}\right) = O\left(\frac{\log^{\frac{1}{2}} r}{\sqrt{K-\frac{M}{N}}}\right) = o\left(C_{\updownarrow}\right)$, thus $C = \Theta\left(\frac{\log^{\frac{3}{2}} r}{\sqrt{K-\frac{M}{N}}}\right)$.
- if $KN-M \overset{\lim}{>} N\ln N$ then $r \sim \frac{KN-M}{\ln N}$. In this case, it is $C_{\downarrow} = O\left(\sqrt{\frac{N}{r}}\right) = O\left(\frac{\log^{\frac{1}{2}} N}{\sqrt{K-\frac{M}{N}}}\right)$. Observe now that the condition of $KN-M \overset{\lim}{\geq} N\ln N$ implies that $r \overset{\lim}{\geq} N$, thus $C = \Theta\left(\frac{\log^{\frac{3}{2}} r}{\sqrt{K-\frac{M}{N}}}\right)$.

Case $\tau >$ 3/2 **and** $KN-M \overset{\lim}{\leq} \frac{2\tau}{2\tau-3}N$: It is $l \to 1$, $r \sim \left[\frac{2\tau-3}{2\tau}(KN-M)\right]^{\frac{3}{2\tau}}$, thus, $C_{\updownarrow} = \Theta\left(\frac{\sqrt{N}}{\sqrt{KN-M}}\right)$. From Lemma 1, for $\mathcal{M}_{\downarrow} \neq \emptyset$, it has to be $M \overset{\lim}{>} \frac{2\tau-3}{2\tau}KN^{\frac{3}{2\tau}}$. Thus, $r \overset{\lim}{\leq} (KN)^{\frac{3}{2\tau}} \overset{\lim}{<} M$, with the last step coming from $\frac{2\tau-3}{2\tau} < 1$. As before, $C_{\downarrow} = \Theta\left(\frac{\sqrt{N}}{r^{\tau-1}}\right) = \Theta\left(\frac{\sqrt{N}}{[KN-M]^{\frac{3}{2}-\frac{3}{2\tau}}}\right) \overset{\tau \geq \frac{3}{2}}{=} o\left(\frac{\sqrt{N}}{\sqrt{KN-M}}\right)$. In total, $C = \Theta\left(\frac{\sqrt{N}}{\sqrt{KN-M}}\right)$.

Case $\tau >$ 3/2 **and** $KN-M \overset{\lim}{>} \frac{2\tau}{2\tau-3}N$: It is $l = \Theta\left(K-\frac{M}{N}\right)$ and $r = \Theta\left([KN-M]N^{\frac{3-2\tau}{2\tau}}\right)$. Thus, $C_{\updownarrow} = \Theta\left(\frac{1}{l^{\tau-\frac{3}{2}}\sqrt{K-\frac{M}{N}}}\right) = \Theta\left(\frac{N^{\tau-1}}{[KN-M]^{\tau-1}}\right)$. Moreover, $C_{\downarrow} = \Theta\left(\frac{\sqrt{N}}{r^{\tau-1}}\right) = \Theta\left(\frac{\sqrt{N}N^{\frac{2\tau-3}{2\tau}(\tau-1)}}{[KN-M]^{\tau-1}}\right) = \Theta\left(\frac{N^{\tau-1}}{[KN-M]^{\tau-1}}\ \frac{1}{N^{\frac{2\tau-3}{2\tau}}}\right) \overset{\tau \geq \frac{3}{2}}{=} o\left(C_{\updownarrow}\right)$.

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

References

[1] D. Trossen, M. Sarela, K. Sollins, Arguments for an information-centric internetworking architecture, ACM SIGCOMM Computer Communication Review 40 (2) (2010) 26–33.

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

[3] K. Katsaros, G. Xylomenos, G. C. Polyzos, A hybrid overlay multicast and caching scheme for information-centric networking, in: Proceedings of 2010 IEEE INFOCOM Conference on Computer Communications Workshops, IEEE, 2010, pp. 1–6.

[4] Cisco, Cisco visual networking index: Global mobile data traffic forecast update, 20102015, Tech. rep. (2011).

[5] P. Gupta, P. R. Kumar, The capacity of wireless networks, IEEE Transactions on Information Theory 46 (2000) 388–404.

[6] S. Gitzenis, G. Paschos, L. Tassiulas, Asymptotic laws for joint content replication and delivery in wireless networks, IEEE Transactions on Information Theory 59 (5) (2013) 2760–2776. doi:10.1109/TIT.2012.2235905.

[7] S. Gitzenis, G. S. Paschos, L. Tassiulas, Asymptotic laws for content replication and delivery in wireless networks, in: Proceedings of the 2012 IEEE INFOCOM conference on Computer Communications, Orlando, FL, USA, 2012, pp. 126–134.

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

[9] G. S. Paschos, S. Gitzenis, L. Tassiulas, The effect of caching in sustainability of large wireless networks, in: Proceedings of the 10th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt 2012),, 2012, pp. 355 –360.

[10] S. Toumpis, Asymptotic capacity bounds for wireless networks with non-uniform traffic patterns, IEEE Transactions on Wireless Communications 7 (2008) 2231–2242.

[11] A. Zemlianov, G. de Veciana, Capacity of ad hoc wireless networks with infrastructure support, IEEE Transactions on Selected Areas of Communications 23 (2005) 657–667.

[12] A. Özgür, O. Lévêque, D. Tse, Hierarchical cooperation achieves optimal capacity scaling in ad hoc networks, IEEE Transactions on Information Theory 53 (2007) 3549–3572.

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

[14] M. Grossglauser, D. N. Tse, Mobility increases the capacity of ad hoc wireless networks, IEEE/ACM Transactions on Networking 10 (4) (2002) 477–486.

[15] M. Garetto, P. Giaccone, E. Leonardi, Capacity scaling in delay tolerant networks with heterogeneous mobile nodes, in: Proceedings of the 8th ACM international symposium on Mobile ad hoc networking and computing, ACM, 2007, pp. 41–50.

[16] E. J. Rosensweig, D. S. Menasche, J. Kurose, On the steady-state of cache networks, in: Proceedings of the IEEE INFOCOM conference on Computer Communications, Torino, Italy, 2013.

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

[18] M. G. G. Alfano, E. Leonardi, Content-centric wireless networks with limited buffers: when mobility hurts, in: Proceedings of 2013 IEEE INFOCOM conference on Computer Communications, Torino, Italy, 2013.

[19] U. Niesen, D. Shah, G. Wornell, Caching in wireless networks, in: Proceedings IEEE International Symposium on Information Theory, Seoul, Korea, 2009, pp. 2111–2115.

[20] J. Silvester, L. Kleinrock, On the capacity of multihop slotted aloha networks with regular structure, IEEE Transactions on Communications 31 (1983) 974–982.

[21] M. Faloutsos, P. Faloutsos, C. Faloutsos, On power-law relationships of the internet topology, in: ACM SIGCOMM Computer Communication Review, Vol. 29, ACM, 1999, pp. 251–262.

[22] J. Munoz-Gea, S. Traverso, E. Leonardi, Modeling and evaluation of multisource streaming strategies in P2P VoD systems, IEEE Transactions on Consumer Electronics 58 (4) (2012) 1202–1210.

[23] V. Lenders, G. Karlsson, M. May, Wireless ad hoc podcasting, in: Proceedings of the 4th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON), 2007, IEEE, 2007, pp. 273–283.

[24] N. Golrezaei, A. G. Dimakis, A. F. Molisch, Wireless device-to-device communications with distributed caching, in: Proceedings of the IEEE International Symposium on Information Theory (ISIT) 2012, IEEE, 2012, pp. 2781–2785.

[25] X. Jiang, P. Gao, Y. Zhao, Y. Shi, Caching scheme research based on unstructured peer-to-peer network, Physics Procedia 25 (2012) 1076–1083.

[26] C. Fricker, P. Robert, J. Roberts, N. Sbihi, Impact of traffic mix on caching performance in a content-centric network, in: Proceedings on the 2012 IEEE INFOCOM conference on Computer Communications Workshops, IEEE, 2012, pp. 310–315.

[27] C. Liaskos, S. Petridou, G. Papadimitriou, P. Nicopolitidis, M. Obaidat, A. Pomportsis, A novel clustering-driven approach to wireless data broadcasting, in: Proceedings of the IEEE/CVR 15th Annual Symposium, 2011.

[28] C. R. Cunha, A. Bestavros, M. E. Crovella, Characteristics of WWW client-based traces, in: View on NCSTRL, Boston University, MA, USA, 1995.

[29] L. Breslau, P. Cue, P. Cao, L. Fan, G. Phillips, S. Shenker, Web caching and Zipf-like distributions: Evidence and implications, in: Proceedings of the 1999 IEEE INFOCOM conference on Computer Communications, New York, NY, USA, 1999, pp. 126–134.

[30] L. A. Adamic, B. A. Huberman, Zipf's law and the Internet, Glottometrics 3 (2002) 143–150.

[31] V. N. Padmanabhan, L. Qiu, The content and access dynamics of a busy web site: Findings and implications, ACM SIGCOMM Computer Communication Review 30 (4) (2000) 111–123.

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

[33] M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, S. Moon, I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system, in: Proceedings of the 7th ACM SIGCOMM conference on Internet measurement, ACM, 2007, pp. 1–14.

[34] G. Dán, N. Carlsson, Power-law revisited: A large scale measurement study of P2P content popularity, in: Proceedings of the International Workshop on Peer-To-Peer Systems (IPTPS), San Jose, CA, USA, 2010.

[35] H. Yu, D. Zheng, B. Y. Zhao, W. Zheng, Understanding user behavior in large-scale video-on-demand systems, in: ACM SIGOPS Operating Systems Review, Vol. 40, ACM, 2006, pp. 333–344.