Computer Networks 64 (2014) 353–368

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

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, 6th km 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*

A R T I C L E I N F O

Article history:
Received 26 May 2013
Received in revised form 5 January 2014
Accepted 6 January 2014
Available online 6 March 2014

Keywords:
Scaling laws
Wireless network capacity
Network sustainability
Content centric networks
Caching
Zipf law

A B S T R A C T

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.

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

* Corresponding author. Address: Savvas Gitzenis, 51 Sapfous Str., 17676 Athens, Greece. Tel.: +30 2109577420.
E-mail addresses: sgitz@iti.gr (S. Gitzenis), gpasxos@gmail.com (G.S. Paschos), leandros@uth.gr (L. Tassiulas).

http://dx.doi.org/10.1016/j.comnet.2014.01.012
1389-1286/

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 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 $O\left(1/\sqrt{N}\right)$, 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 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 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 $O\left(1/\sqrt{N}\right)$ 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

[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].