Cache Increases the Capacity of Wireless Networks

Li Qiu and Guohong Cao
Department of Computer Science and Engineering
The Pennsylvania State University
E-mail: {lyq5023, gcao}@cse.psu.edu

***Abstract*—Caching in wireless ad hoc networks can reduce network traffic and content access delay, as nodes can retrieve contents from near neighbors rather than the faraway server. However, the fundamental performance limits of caching in wireless ad hoc networks have rarely been studied in an analytical manner. In this paper, we study the fundamental property of wireless networks with caching, i.e., the scaling laws of the network capacity based on cache size of individual node, the total size of unique content and the number of nodes in the network. We present an upper bound on network capacity, and present an achievable capacity lower bound, where we propose a caching scheme to show what capacity can actually be achievable. Our results suggest that the capacity of wireless ad hoc networks with caching can remain constant even as the number of nodes in the network increases. We also present numerical results and demonstrate that our results are consistent with existing analytical results under extreme conditions where the node communication scenario matches theirs.**

I. Introduction

Over the past few years, caching in wireless ad hoc networks has attracted lots of attention. By caching contents at multiple nodes, network traffic and content access delay can be reduced as users can access contents from near neighbors instead of the faraway server. Although various caching techniques [1]–[4] have been proposed to improve the performance of wireless ad hoc networks, these studies are limited to protocol design and the associated performance evaluations are based on simulations. The fundamental performance limits of caching in wireless ad hoc networks have rarely been studied in an analytical manner. Although there are some analytic performance studies on caching in Internet such as modeling the cache hit ratio [5]–[7], modeling the steady-states of cache networks [8], performance analysis of optimal routing and content caching [9], etc., these results cannot be directly applied to wireless ad hoc networks.

There are some well known analytical results related to the capacity of wireless ad hoc networks, which is constrained by the mutual interference of concurrent transmissions between nodes. Gupta and Kumar [10] have proved that each node can transmit at most $\Theta(\frac{W}{\sqrt{n}})$ bits per second, where n is the number of nodes and W is the channel throughput. That is, the network capacity decreases as the number of nodes increases. Later, Grossglauser and Tse [11] show that the per-user throughput can increase dramatically when nodes are mobile rather than static. It is possible for network capacity to remain constant even if the number of nodes increases, but at the cost of long data transmission delay. However, cache is not considered in their studies, and it is not clear how cache affects the capacity of wireless ad hoc networks.

In this paper, we study the capacity of wireless ad hoc networks with caching with respect to the cache size of each node (s), the total size of unique content in the network (m), and the number of nodes (n) in the network. As various caching schemes may lead to totally different performance, we first investigate how the employed caching scheme will affect the network capacity. Based on the effect of the caching schemes, we derive a capacity upper bound for caching in wireless ad hoc networks. However, this upper bound only specifies the maximum capacity that the network could possibly support, instead of the capacity that is actually achievable. To address this problem, we also design a caching scheme and show that by carefully caching distinct contents at specific nodes, a capacity of $\Theta\left(W\frac{s}{m}\right)$ can always be achievable. The major contributions of the paper are summarized as follows:

- We find an upper bound on the capacity of wireless ad hoc networks with caching.
- We are the first to find an achievable capacity lower bound for wireless ad hoc networks with caching.
- We present numerical results and demonstrate that our results are consistent with existing analytical results under extreme conditions where the node communication scenario matches theirs.
- Our analytical results suggest that for wireless ad hoc networks with caching, the network capacity can remain constant even if the number of nodes increases. This is in sharp contrast to previous results on capacity of wireless ad hoc networks without caching, in which the network capacity decreases quickly as the number of nodes increases [10].

The rest of the paper is organized as follows. We review a few related works in Section II. We give the models and definitions in Section III. In Section IV, we derive a capacity upper bound for wireless networks with caching. We construct an achievable capacity lower bound in Section V. In Section VI, we present the numerical results and discuss their implications. We conclude the paper in Section VII.

II. Related Work

Caching in wireless networks has been studied from various aspects. Yin $et\ al.$ [1], [2] have examined several cooperative caching schemes to determine where to cache the data. Jin and

This work was supported in part by the National Science Foundation (NSF) under grant number CNS-1320278 and CNS-1526425.

Wang [12] have proposed solutions to determine the optimal placement of replicas in the network. Fiore *et al*. [3] have proposed techniques to determine whether a node should cache the data to reduce data redundancy among neighbors. However, none of them have studied the fundamental performance limits of caching in wireless networks in an analytical manner.

Niesen *et al*. [13] studied the content delivery problem in wireless networks from an information-theoretical point of view. They focused on deriving the region of feasible request serving rate by knowing where and what data has been cached, while we focus on the scaling laws of network capacity based on the cache size and the number of nodes in the network.

In [14], Gitzenis *et al.* studied the asymptotic laws for joint replication and delivery in wireless networks. They derived the minimum throughput on each link so that every node is able to satisfy one request per second. However, in practice, the transmissions on various links may interfere with each other. Thus it is unknown whether their throughput can be supported by the network, and it is unclear what is the network capacity actually achievable. Furthermore, we present more interesting results of caching which show that the network capacity can remain constant even if the number of nodes increases.

Azimdoost *et al.* [15] studied the capacity of wireless networks with caching when each content has a limited lifetime. They presented a capacity upper bound for both grid and random networks. In [16], the authors derived capacity upper bounds for two specific content access schemes considering the number of nodes n and the cache size s. However, in both [15] and [16], the authors only consider the capacity upper bound, but fail to investigate what capacity is actually achievable. In addition, both assume that the wireless transmission range is in the order of $\sqrt{\frac{\log n}{n}}$, which will significantly restrict the network capacity.

III. MODEL

A. Network Model

We consider a wireless ad hoc network where n nodes are independently and uniformly distributed on the surface of a unit sphere. Similar to [10], we analyze the capacity of wireless networks with caching on the surface of the sphere S^2 rather than on a disk so as to eliminate the edge effects; i.e., nodes near the edge have much fewer neighbors than nodes near the center.

Assume the nodes are homogeneous, i.e., each node can cache s bits of contents, and all transmissions employ the same amount of power P. All nodes transmit on a common wireless channel which can support W bits per second. Let $\tau(t)$ denote the set of nodes simultaneously transmitting at time t. Suppose a node $i \in \tau(t)$ sends data to node j, according to [10], the transmission rate can reach W bits per second if:

$$\frac{\frac{P}{X_{i,j}^{\alpha}}}{N_0 + \sum_{\substack{k \in \tau(t) \\ k \neq i}} \frac{P}{X_{k,j}^{\alpha}}} \geq \beta, \tag{1}$$

where β is the minimum SIR for successful reception, $X_{i,j}$ is the great-circle distance (i.e., the shortest distance between two points on the surface of a sphere) between i and j, N_0 is the white noise and α is a parameter larger than 2 describing how the signal strength scales with distance.

B. Content Access Model

Let $\Phi = \{\rho_i\}_{1 \leq i \leq m}$ denote the set of m unique content throughout the network. To simplify the analysis, we assume each content has one bit. Note that our capacity analysis results can be applied to cases where the content has various sizes, as our analysis only depends on the total content size rather than the size of individual content. These contents are cached throughout the network where each node i caches a subset ϕ_i of Φ locally. The cache size constraint requires $|\phi_i| \leq s$.

When each content has one bit, m contents have m bits. Apparently, when $s \geq m$, the problem is trivial since each node can cache all m contents and then all content requests can be satisfied by local cache. On the other hand, to guarantee that at least one copy of each content exists in the network, the total cache size should be larger than m, i.e., $ns \geq m$. Thus, we assume $\frac{m}{n} \leq s < m$.

At each node, there is always a content request, and a new request will arrive after the previous request has been served. The content requests follow uniform distribution. If the requested content has been cached locally, it will be directly served. Otherwise, the node receiving the request has to contact other nodes for the content either directly or through multiple hops.

C. Capacity and Request Satisfaction Rate

Similar to [10], [11], network capacity describes node's capability to transmit or retrieve contents. Under our content access model, the capacity of wireless networks $\lambda(s, m, n)$ is the number of bits that nodes can receive from others per second.

In addition to capacity, we also analyze the request satisfaction rate $\mu(s, m, n)$, which is the number of requests that can be satisfied per second under our content access model. As the request satisfaction rate depends on both local cache and network capacity, it provides a more complete description of how the cache size may affect the caching performance in wireless networks.

The achievable capacity and request satisfaction rate may depend on the location of nodes. As the nodes are uniformly distributed, there is always a small probability for them to be extremely oddly located, and then the capacity and the request satisfaction rate may be largely deviated. Therefore, when defining the upper and lower bounds, we allow for vanishingly small probability that the actual value deviates from the upper or lower bounds.

Definition 1 (Upper Bound)**.** *The capacity (request satisfaction rate) is upper bounded by $\hat{\lambda}(s, m, n)$ ($\hat{\mu}(s, m, n)$), if there exists a constant c (c') such that for any caching scheme and transmission schedule*

$$\begin{aligned} &\lim_{n \to \infty} \Pr\left(\lambda(s, m, n) \leq c\hat{\lambda}(s, m, n)\right) = 1 \\ &\lim_{n \to \infty} \Pr\left(\mu(s, m, n) \leq c'\hat{\mu}(s, m, n)\right) = 1. \end{aligned} \tag{2}$$

Definition 2 (Achievable Lower Bound)**.** *A capacity (request satisfaction rate) of* $\tilde{\lambda}(s,m,n)$ *(*$\tilde{\mu}(s,m,n)$*) is achievable, if there exists a constant* c *(*c'*), and a caching scheme and transmission schedule, such that*

$$\begin{aligned} &\lim_{n\to\infty} \Pr\left(\lambda(s,m,n) \geq c\tilde{\lambda}(s,m,n)\right) = 1 \\ &\lim_{n\to\infty} \Pr\left(\mu(s,m,n) \geq c'\tilde{\mu}(s,m,n)\right) = 1. \end{aligned} \tag{3}$$

IV. An Upper Bound on Network Capacity

In this section, we present an upper bound on the capacity of wireless networks with caching. We first show that nodes prefer the contents cached at closest nodes to achieve a higher capacity. Then we derive the network capacity for an ideal scenario where each node has contents cached as close as possible, which will be the upper bound.

A. Capacity and Transmission Distance

For a node i, after receiving requests for content ρ_j not cached locally, it has to retrieve ρ_j from the node that caches the content. Under different caching schemes, i may contact nearby nodes or faraway nodes to retrieve ρ_j, and then lead to different performance. To derive a capacity upper bound for all possible schemes, we first discuss what may affect the capacity.

In a wireless network with n nodes, where each node transmits to an arbitrarily chosen destination, Gupta and Kumar [10] prove that the per node capacity is upper bounded by $\frac{W}{L\sqrt{n}}$, where L is the average distance between the source and destination pairs. This result indicates that as nodes communicate with more distant nodes, the capacity decreases. A similar result has also been found in [11], [17], where the authors suggest that the capacity can be improved by shortening the transmission distance between the source node and the destination node. This is because as the transmission distance increases, more intermediate nodes will be affected and then less number of simultaneous transmissions are allowed.

Based on the above results, each node prefers the other $m-s$ contents (s contents can be cached locally) to be cached at nodes as close as possible to achieve higher capacity. Thus, to find an upper bound, we assume each node i caches s contents locally, and the closest $\lceil \frac{m-s}{s} \rceil$ nodes cache the remaining $m-s$ contents. Let $\{X^i_{(1)}, ..., X^i_{(n-1)}\}$ denote the ordered distance between i and any other node, i.e., $X^i_{(j)}$ is the distance between i and its j-th closest neighbor. For simplicity, we also use $X^i_{(j)}$ to represent the j-th closest neighbor of i. Then i and $X^i_{(1)}, ..., X^i_{(\lceil \frac{m}{s}-2\rceil)}$ will each cache s distinct contents, and $X^i_{(\lceil \frac{m}{s}-1\rceil)}$ will cache the remaining $m - s\lceil \frac{m}{s} - 1\rceil$ contents. Although such an ideal scenario where every node has the remaining contents cached at its closest neighbors may not always be practical, there may be special cases when such requirement can be satisfied, and thus this ideal scenario can be used for calculating the capacity upper bound.

B. Minimum Expected Transmission Distance

In this subsection, we focus on deriving the average data transmission distance in the aforementioned ideal scenario. As all nodes are uniformly and independently distributed on the surface of the sphere with radius 1, given location of node i, the distance between i and any other node k follows the distribution shown below:

$$\begin{aligned} f(x) &= \frac{\sin x}{2}(0 \leq x \leq \pi), \\ F(x) &= \frac{1-\cos x}{2}(0 \leq x \leq \pi), \end{aligned} \tag{4}$$

where $f(x)$ and $F(x)$ are the pdf and cdf, respectively.

Node i has $n-1$ neighbors, and the distance between i and its neighbors will be $n-1$ i.i.d random variables following the above distribution. Therefore, the ordered random variables $\{X^i_{(1)}, ..., X^i_{(n-1)}\}$ are equivalent to $n-1$ ordered random variables drawn from distribution (4).

For i's neighbors $X^i_{(j)}$ $(1 \leq j \leq \lceil \frac{m}{s} - 2\rceil)$, i will retrieve contents from any of them with equivalent probability $\frac{s}{m-s}$ as each of them caches s distinct contents. Specifically, $X^i_{(\lceil \frac{m}{s}-1\rceil)}$ may cache fewer useful contents for i, and i might retrieve contents from it less frequently. Let $\xi = \lceil \frac{m}{s}\rceil - 1$, and let L_i denote the average transmission distance of the contents sent to i, then the expectation of L_i will be

$$\mathrm{E}(L_i) = \mathrm{E}\left(\frac{s\sum_{j=1}^{\xi-1} X^i_{(j)} + (m-s\xi)\, X^i_{(\xi)}}{m-s}\right), \tag{5}$$

where $X^i_{(j)}$ is the distance between i and its j-th closest neighbor. Based on Theorem 2.5 of [18], given $X^i_{(r)} = x_r$ and $X^i_{(s)} = x_s$ $(r < s)$, the conditional distribution of $X^i_{(r+1)}$, ..., $X^i_{(s-1)}$ is the distribution of $s-r-1$ random variables drawn from $f(x)/[F(x_s) - F(x_r)]$ $(x_r \leq x \leq x_s)$. Thus, given $X^i_{(\xi)}$, variables $X^i_{(1)}$, ..., $X^i_{(\xi-1)}$ are equivalent to $\xi - 1$ random variables drawn from the distribution of

$$g_{X^i_{(\xi)}}(x) = \frac{f(x)}{F\left(X^i_{(\xi)}\right)}\left(0 \leq x \leq X^i_{(\xi)}\right). \tag{6}$$

Suppose a random variable $Y^i_{(\xi)}$ follows the distribution $g_{X^i_{(\xi)}}$. Then, the expectation of $Y^i_{(\xi)}$ is

$$\begin{aligned} \mathrm{E}(Y^i_{(\xi)}) &= \mathrm{E}(\mathrm{E}(Y^i_{(\xi)}|X^i_{(\xi)})) = \mathrm{E}\left(\int_0^{X^i_{(\xi)}} g_{X^i_{(\xi)}}(x)\, x\, \mathrm{d}x\right) \\ &= \mathrm{E}\left(\frac{2}{3}X^i_{(\xi)} - \frac{1}{90}{X^i_{(\xi)}}^3 + \ldots\right), \end{aligned}$$

where the result on the last line is the Taylor series of the integral on the second line. To simplify the results, we approximate $\mathrm{E}(Y^i_{(\xi)})$ to the first term of the Taylor series, which leads to

$$\mathrm{E}(Y^i_{(\xi)}) \approx \frac{2}{3}\mathrm{E}(X^i_{(\xi)}). \tag{7}$$

Note that $\mathrm{E}(X^i_{(\xi)})$ is smaller than π, thus the truncation error is insignificant (see simulation results). Combining formula (5) with formula (7), we can get

$$\begin{aligned} \mathrm{E}(L_i) &= \frac{s(\xi-1)}{m-s}\mathrm{E}\left(Y^i_{(\xi)}\right) + \frac{m-s\xi}{m-s}\mathrm{E}\left(X^i_{(\xi)}\right) \\ &\approx \frac{2s(\xi-1)}{3(m-s)}\mathrm{E}\left(X^i_{(\xi)}\right) + \frac{m-s\xi}{m-s}\mathrm{E}\left(X^i_{(\xi)}\right) \\ &\geq \frac{2}{3}\mathrm{E}\left(X^i_{(\xi)}\right). \end{aligned} \tag{8}$$

Unfortunately, $\mathrm{E}(X^i_{(\xi)})$ is not easy to compute. As $X^i_{(\xi)}$ is the ξ-th smallest among $n-1$ random variables drawn from distribution (4), $X^i_{(\xi)}$ follows the distribution of

$$\begin{aligned} f_{X^i_{(\xi)}}(x) = & \binom{n-1}{\xi-1} (F(x))^{\xi-1} \\ & \times \binom{n-\xi}{n-\xi-1} (1-F(x))^{n-\xi-1} f(x). \end{aligned}$$

The expectation of the random variable is

$$\mathrm{E}\left(X^i_{(\xi)}\right) = \int_0^\pi x f_{X^i_{(\xi)}}(x)\,\mathrm{d}x. \tag{9}$$

The above expectation is extremely difficult to obtain. The exact expectation of a similar distribution can be found in [19], where the only difference is that they consider a distribution defined in $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, while we consider a distribution defined in $[0,\pi]$. The exact expression is presented in theorem 2.1 of their paper which contains both Beta function and Gamma function, and is apparently too evolved for later analysis. Thus, we exploit the David and Johnson series to approximate $\mathrm{E}\left(X^i_{(\xi)}\right)$ [18]. David and Johnson state that given cdf $F(x)$, the expected value of $X^i_{(\xi)}$ can be approximated by:

$$\mathrm{E}\left(X^i_{(\xi)}\right) = Q(c_\xi) + \frac{c_\xi d_\xi}{2(n+1)} Q''(c_\xi) + O\left(\frac{1}{(n+1)^2}\right),$$

where $Q = F^{-1}$, $c_\xi = \frac{\xi}{n}$, and $d_\xi = 1 - c_\xi$. As $F(x) = \frac{1}{2}(1-\cos x)$, then

$$\begin{aligned} Q(x) &= F^{-1}(x) = \arccos(1-2x) \\ Q'(x) &= \frac{\mathrm{d}Q}{\mathrm{d}x} = \frac{1}{\sqrt{x-x^2}} \\ Q''(x) &= \frac{\mathrm{d}^2Q}{\mathrm{d}x^2} = \frac{2x-1}{2(x-x^2)^{-\frac{3}{2}}}. \end{aligned}$$

Then the expectation of $X^i_{(\xi)}$ is

$$\mathrm{E}(X^i_{(\xi)}) = \arccos(1-\frac{2\xi}{n}) + \frac{1-\frac{2\xi}{n}}{2(n+1)\sqrt{\frac{\xi}{n}-\frac{\xi^2}{n^2}}} + \dots$$

For simplicity, we approximate $\mathrm{E}(X^i_{(\xi)})$ to its first term of David and Johnson series:

$$\mathrm{E}(X^i_{(\xi)}) \approx \arccos(1-\frac{2\xi}{n}). \tag{10}$$

Note that the truncation error is insignificant. Even in the extreme case of $\xi = 1$ (i.e., when the David and Johnson series is most prone to approximation error), the exact value of $\mathrm{E}(X^i_{(1)}) = \mathrm{Beta}(n-\frac{1}{2}, \frac{1}{2})$ [20] which is $\frac{\sqrt{\pi}}{\sqrt{n}}$ when n is large (based on Stirling's approximation), while our approximated result is $\frac{2}{\sqrt{n}}$. These two only differ by a small constant factor, and is already good enough for our analysis. We will also use simulations to verify this approximation later.

Combining formula (8) with formula (10), we can get

$$\mathrm{E}(L_i) \geq \frac{2}{3}\arccos\left(1-\frac{2\xi}{n}\right). \tag{11}$$

Note that i is a node arbitrarily chosen, thus the above inequality actually applies to all nodes in the network.

C. Capacity Upper Bound

Recall L_i is the average transmission distance of contents sent to i, the average transmission distance for all contents will be

$$\bar{L} = \frac{\sum_{i=1}^n L_i}{n}. \tag{12}$$

Then, $\mathrm{E}\left(\bar{L}\right) = \mathrm{E}(L_i)$, as $L_1, L_2, L_3, ..., L_n$ are identically distributed.

Lemma 1. *The average transmission distance $\bar{L}$ converges in probability to $\mathrm{E}\left(\bar{L}\right)$ as $n \to \infty$, i.e., for any $\epsilon > 0$,*

$$\lim_{n\to\infty} \Pr\left(|\bar{L} - \mathrm{E}\left(\bar{L}\right)| \geq \epsilon\right) = 0.$$

Proof. Based on formula (5),

$$\mathrm{Var}(L_i) = \mathrm{Var}\left(\frac{s\sum_{j=1}^{\xi-1} X^i_{(j)} + (m-s\xi)X^i_{(\xi)}}{m-s}\right).$$

Let $a_1 = a_2 = ... = a_{\xi-1} = \frac{s}{m-s}$, and $a_\xi = \frac{m-s\xi}{m-s}$, then

$$\begin{aligned} \mathrm{Var}(L_i) &= \mathrm{Var}\left(\sum_{j=1}^{\xi} a_j X^i_{(j)}\right) \\ &= \sum_{j=1}^{\xi} \mathrm{Var}\left(a_j X^i_{(j)}\right) + \sum_{j=1}^{\xi}\sum_{\substack{k=1\\k\neq j}}^{\xi} \mathrm{Cov}\left(a_j X^i_{(j)}, a_k X^i_{(k)}\right). \end{aligned}$$

Based on David and Johnson series [18],

$$\mathrm{Var}\left(X^i_{(j)}\right) = \frac{c_j d_j}{n+1} Q'(c_j)^2 + O\left(\frac{1}{n^2}\right).$$

Recall that $Q'(c_j) = \frac{1}{\sqrt{c_j - c_j^2}}$ and $d_j = 1 - c_j$, therefore for any i, j

$$\mathrm{Var}\left(X^i_{(j)}\right) = \frac{1}{n+1} + O\left(\frac{1}{n^2}\right) = O\left(\frac{1}{n}\right).$$

In addition, for any i, j and k

$$\begin{aligned} \mathrm{Cov}\left(X^i_{(j)}, X^i_{(k)}\right) &\leq \sqrt{\mathrm{Var}\left(X^i_{(j)}\right)\mathrm{Var}\left(X^i_{(k)}\right)} \\ &\leq \max\left(\mathrm{Var}\left(X^i_{(j)}\right), \mathrm{Var}\left(X^i_{(k)}\right)\right). \end{aligned}$$

In this way, the variance of L_i is bounded by

$$\begin{aligned} \mathrm{Var}(L_i) &\leq \left(\sum_{j=1}^{\xi} a_j^2 + \sum_{j=1}^{\xi}\sum_{\substack{k=1\\k\neq j}}^{\xi} a_j a_k\right) \max_{1\leq j\leq\xi}\left(\mathrm{Var}\left(X^i_{(j)}\right)\right) \\ &= \left(\sum_{j=1}^{\xi} a_j\right)^2 \max_{1\leq j\leq\xi}\left(\mathrm{Var}\left(X^i_{(j)}\right)\right) = O\left(\frac{1}{n}\right). \end{aligned}$$

The above formula holds for all i, thus the variance of $\bar{L}$ is

$$\begin{aligned} \mathrm{Var}\left(\bar{L}\right) &= \frac{1}{n^2}\left(\sum_{i=1}^{n}\mathrm{Var}(L_i) + \sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n}\mathrm{Cov}(L_i, L_j)\right) \\ &\leq \frac{n+n(n-1)}{n^2}\max_{1\leq i\leq n}(\mathrm{Var}(L_i)) = O\left(\frac{1}{n}\right). \end{aligned}$$

Based on Chebyshev's inequality [21], for any $\epsilon > 0$,

$$P\left(\left|\bar{L} - \mathrm{E}\left(\bar{L}\right)\right| \geq \epsilon\right) \leq \frac{Var\left(\bar{L}\right)}{\epsilon^2} \leq \frac{1}{n\epsilon^2}.$$

As $n \to \infty$, the above probability goes to 0. □

Theorem 1. *Under our network and content access model, the per node capacity is upper bounded by*

$$\hat{\lambda}(s, m, n) = W\sqrt{\frac{s}{m}}. \tag{13}$$

Proof. In the proof of Theorem 5.2 [10], the authors identified that for nodes uniformly distributed on the surface of the sphere, if the average transmission distance is $\bar{L}$, then the capacity for each node can not be larger than $\frac{\sqrt{8}}{\sqrt{\pi}} \frac{W}{\beta^{\frac{1}{\alpha}}-1} \frac{1}{\bar{L}\sqrt{n}}$. We have shown that even in the ideal scenario where each node retrieves contents from closest neighbors, the average transmission distance $\bar{L} \xrightarrow{p} \mathrm{E}(\bar{L}) \geq \frac{2}{3} \arccos\left(1 - \frac{2\xi}{n}\right)$, thus per node capacity

$$\begin{aligned}
\lambda(s, m, n) &\leq \frac{\sqrt{8}}{\sqrt{\pi}} \frac{W}{\beta^{\frac{1}{\alpha}} - 1} \frac{1}{\bar{L}\sqrt{n}} \\
&\leq \frac{\sqrt{8}}{\sqrt{\pi}} \frac{W}{\beta^{\frac{1}{\alpha}} - 1} \frac{1}{\sqrt{n}} \frac{3}{2\arccos(1 - \frac{2}{n}(\lceil \frac{m}{s} \rceil - 1))} \\
&= O\left(W\sqrt{\frac{s}{m}}\right) \\
&\text{(as } \left\lceil \frac{m}{s} \right\rceil \geq 2 \text{ and } \arccos(1 - 2x) \approx 2\sqrt{x}).
\end{aligned}$$

□

Corollary 1.1. *Under our network and content access model, the per node request satisfaction rate is upper bounded by*

$$\hat{\mu}(s, m, n) = W\sqrt{\frac{s}{m}} \frac{m}{m - s}. \tag{14}$$

Proof. Each node caches s out of m contents locally. Since the content requests follow the uniform distribution, at each node, $\frac{s}{m}$ of all requests are served by local cache, while the remaining $\frac{m-s}{m}$ of requests are served by other nodes. It has been shown that no more than $\hat{\lambda}(s, m, n)$ bits can be received at every node per second. Thus no more than $\hat{\lambda}(s, m, n)$ requests where the requested contents not in local cache can be satisfied per second. Thus, the per node request satisfaction rate is upper bounded by $\frac{m}{m-s}\hat{\lambda}(s, m, n)$ (note that requests come one after another). Therefore,

$$\hat{\mu}(s, m, n) = \frac{m}{m - s}\hat{\lambda}(s, m, n) = W\sqrt{\frac{s}{m}} \frac{m}{m - s}.$$

□

V. Achievable Capacity Lower Bound

In this section, we construct a caching scheme to show that at least a capacity of $\Theta\left(W\frac{s}{m}\right)$ is achievable. Basically, we partition the surface of the sphere S^2 into cells with similar sizes. Then we show that by carefully choosing the contents to cache at each cell, most of the nodes can retrieve any content from nodes within the same cell. Additionally, we prove that transmission within the same cell will only affect a constant number of neighboring cells. Based on the above results, we prove that a capacity proportional to the number of cells can be achieved.

A. Voronoi Tessellation

In our constructed caching scheme, to partition the sphere into cells, we use Voronoi Tessellation [22], so that the divided cells have similar area. Given the t initial points $\{b_1, b_2, ..., b_t\}$, Voronoi cell V_i $(1 \leq i \leq t)$ consists of all the points that are closer to b_i than to any other b_j:

$$V_i = \left\{x \in S^2 \middle| \, |x - b_i| \leq |x - b_j| \text{ for all } 1 \leq j \leq t\right\},$$

where $|x - b_i|$ is the great circle distance between x and b_i, and x can be any point on the surface of the sphere.

As shown in Lemma 4.1 of [10], for any $\rho > 0$, there is a Voronoi tessellation which can partition the sphere into cells such that each cell contains a disk of radius ρ and the cell is contained in a disk of radius 2ρ. With such a property of Voronoi tessellation, we set ρ as the following:

$$\rho := \text{radius of the disk with area } \frac{8\pi}{n}\left\lceil\frac{m}{s}\right\rceil \text{ on } S^2. \tag{15}$$

Since each Voronoi cell contains a disk of radius ρ, and it is contained in a disk of radius 2ρ, its area is at least $\frac{8\pi}{n}\left\lceil\frac{m}{s}\right\rceil$, and at most $c_1\frac{32\pi}{n}\left\lceil\frac{m}{s}\right\rceil$, where c_1 is a constant close to 1, since the area of a disk with radius 2ρ on the surface of the sphere is different from $4\pi\rho^2$. As the surface of the sphere has an area of 4π, the number of cells t must satisfy

$$\begin{aligned}
4\pi/c_1\frac{32\pi}{n}\left\lceil\frac{m}{s}\right\rceil \leq t \leq 4\pi/\frac{8\pi}{n}\left\lceil\frac{m}{s}\right\rceil \\
\frac{n}{8c_1\left\lceil\frac{m}{s}\right\rceil} \leq t \leq \frac{n}{2\left\lceil\frac{m}{s}\right\rceil}.
\end{aligned} \tag{16}$$

For the special case when $\left\lceil\frac{m}{s}\right\rceil \geq \frac{n}{4}$, we have the surface of the sphere as one cell and it contains all n nodes. The number of cells $t = 1$, and it is still in the order of $n/\left\lceil\frac{m}{s}\right\rceil$.

B. Finding cells that can cache all content

In our caching scheme, we want to make sure that most nodes can retrieve content from nodes in the same cell, which means that nodes in the cell can cache all content. Thus, we want to find the cells which contain at least $\left\lceil\frac{m}{s}\right\rceil$ nodes, since only these cells can cache all content. To simplify the notation, let $\eta = \left\lceil\frac{m}{s}\right\rceil$.

Lemma 2. *For any cell V_i $(1 \leq i \leq t)$,*

$$\Pr\left(V_i \text{ contains at least } \eta \text{ nodes}\right) \geq 1 - e^{-\frac{1}{4}}. \tag{17}$$

Proof. As nodes are uniformly distributed, and V_i has an area of at least $\frac{8\pi}{n}\eta$, the probability p_i for any node j to fall into V_i is:

$$\Pr\left(\text{Node } j \text{ falls in } V_i\right) = p_i \geq \frac{8\pi}{n}\eta/4\pi = \frac{2}{n}\eta.$$

Let v_i denote the number of nodes in cell V_i, as n nodes are independently distributed, v_i follows a binomial distribution

$\mathrm{B}(n, p_i)$. For the cumulative probability $\Pr(v_i \le v)$, Chernoff bound [23] states that the probability for variable v_i to be smaller than a constant v is bounded by

$$\Pr(v_i \le v) \le \exp\left(-\frac{(np_i - v)^2}{2np_i}\right), \text{ for any } v \le np_i.$$

By setting $v = \eta$ (note $v < 2\eta \le np_i$), we can get

$$\begin{aligned}\Pr(v_i \le \eta) &\le \exp\left(-\frac{(np_i - \eta)^2}{2np_i}\right) \\ &\le \exp\left(-\frac{(np_i - \frac{np_i}{2})^2}{2np_i}\right) = \exp\left(-\frac{np_i}{8}\right) \\ &\le \exp\left(-\frac{\eta}{4}\right) \le \exp\left(-\frac{1}{4}\right).\end{aligned}$$

Then the probability that V_i contains at least η nodes is

$$\Pr(v_i \ge \eta) \ge 1 - \Pr(v_i \le \eta) \ge 1 - e^{-\frac{1}{4}}.$$

As we have not made any assumption on cell V_i, the above inequality actually holds for all cells. □

Let Bernoulli random variables $\{u_i\}_{1 \le i \le t}$ denote if each cell has more than η nodes (i.e., $u_i = 1$ if $v_i \ge \eta$, and $u_i = 0$ otherwise). Let $\sigma = (\sum_{i=1}^t u_i)/t$, then

$$\mathrm{E}(\sigma) = \frac{\sum_{i=1}^t \Pr(v_i \ge \eta)}{t} \ge 1 - e^{-\frac{1}{4}}. \tag{18}$$

Lemma 3. *For any $\epsilon > 0$,*

$$\lim_{n \to \infty} \Pr(|\sigma - \mathrm{E}(\sigma)| \ge \epsilon) = 0. \tag{19}$$

Proof. There are two cases.

(i) $\eta \le 50 \log n$:

For any u_i and u_j,

$$\begin{aligned}\mathrm{Cov}(u_i, u_j) &= \mathrm{E}(u_i u_j) - \mathrm{E}(u_i)\mathrm{E}(u_j) \\ &= \Pr(u_j = 1 | u_i = 1)\Pr(u_i = 1) \\ &\quad - \Pr(u_j = 1)\Pr(u_i = 1).\end{aligned}$$

Conditional on $u_i = 1$, the probability $\Pr(u_j = 1 | u_i = 1)$ will not be as large as $\Pr(u_j = 1)$. Because $u_i = 1$ implies that at least η nodes are known to fall into V_i, and fewer nodes might fall into V_j. Thus, the above covariance

$$\begin{aligned}\mathrm{Cov}(u_i, u_j) &= (\Pr(u_j = 1 | u_i = 1) - \Pr(u_j = 1)) \\ &\quad \times \Pr(u_i = 1) \le 0.\end{aligned}$$

Therefore, the variance of σ is

$$\begin{aligned}\mathrm{Var}(\sigma) &= \mathrm{Var}\left(\frac{\sum_{i=1}^t u_i}{t}\right) \\ &= \frac{\sum_{i=1}^t \mathrm{Var}(u_i) + \sum_{i=1}^t \sum_{\substack{j=1 \\ j \ne i}}^t \mathrm{Cov}(u_i, u_j)}{t^2} \\ &\le \frac{\sum_{i=1}^t \mathrm{Var}(u_i)}{t^2} \le \frac{\sum_{i=1}^t \Pr(u_i = 1)}{t^2} \\ &\le \frac{1}{t}.\end{aligned}$$

Based on Chebyshev's inequality, $\forall \epsilon > 0$,

$$\Pr(|\sigma - \mathrm{E}(\sigma)| \ge \epsilon) \le \frac{\mathrm{Var}(\sigma)}{\epsilon^2} \le \frac{1}{t\epsilon^2}.$$

Note that $t = \Theta(n/\eta)$ and $\eta \le 50 \log n$, thus $t \ge c_2 \frac{n}{\log n}$, and above probability approaches 0 as n goes to infinity.

(ii) $\eta > 50 \log n$:

If $\eta \ge \frac{n}{4}$, there is only one cell and it contains all n nodes. As $n \ge \eta$, $\Pr(\sigma = 1) = 1$.

Otherwise, based on *Vapnik-Chervonenkis Theorem* [24], for a set of subsets U with finite VC-dimension, and n i.i.d. random variables $\{X_i\}$ with common probability distribution G,

$$\Pr\left(\max_{D_i \in U} \left|\frac{I(D_i)}{n} - G(D_i)\right| \le \epsilon\right) > 1 - \delta$$

if n satisfies

$$n > \max\left\{\frac{VC - d(U)}{\epsilon} \log \frac{16e}{\epsilon}, \frac{4}{\epsilon} \log \frac{2}{\delta}\right\}.$$

Here $I(D_i)$ is the number of random variables that falls in D_i, and $VC - d(U)$ is the VC-dimension of U. Based on Lemma 4.7 in [10], the VC-dimension of the set of disks on S^2 strictly smaller than hemisphere is 3. Thus, if we let U be the set of disks with radius ρ on S^2, and

$$\epsilon = \frac{\eta}{n}, \text{ and } \delta = \frac{50 \log n}{n},$$

then the following inequality is always satisfied:

$$\Pr\left(\max_{D_i \in U} \left|\frac{I(D_i)}{n} - \frac{2\eta}{n}\right| \le \frac{\eta}{n}\right) > 1 - \frac{50 \log n}{n}.$$

It can be rewritten as

$$\Pr(\text{for every } D_i \in U, I(D_i) \ge \eta) > 1 - \frac{50 \log n}{n}. \tag{20}$$

As each cell V_i contains a disk of radius ρ,

$$\begin{aligned}\Pr(v_i \ge \eta, \text{for every } 1 \le i \le t) > 1 - \frac{50 \log n}{n} \\ \Pr(\sigma = 1) > 1 - \frac{50 \log n}{n}.\end{aligned}$$

□

The above lemma states that most of the cells can cache all contents when ρ is set up based on formula (15).

C. Transmission Scheduling

In this subsection, we propose a transmission schedule which shows how the contents can be retrieved in our caching scheme. We show that under the proposed schedule, for a time period of T, cWT can be transmitted within each cell where c is a constant. A similar fact with slightly different settings has already been established in Lemma 4.4 of [10], yet we still include the proof here for completeness. In what follows, we assume that at most one node is transmitting data at any time in each cell.

Definition 3 (Interfering Neighbors)**.** *Two cells are interfering neighbors if there is one point in one cell and another point in another cell that is at most $\delta\rho$ away.*

Lemma 4. *For sufficiently large transmission power P, there exists a constant δ such that if there are no concurrent transmissions from two interfering neighbors, any transmission between nodes within the same cell can always be successfully received.*

Proof. Consider a node i transmitting to a node j in the same cell. As each cell must be contained in a disk of radius 2ρ, the transmission distance between i and j is at most 4ρ. The signal power received at node j is at least $\frac{P}{(4\rho)^{\alpha}}$.

Since there is no concurrent transmissions from interfering neighbors, any two nodes transmitting simultaneously must be separated by a distance of at least $\delta\rho$. Therefore, disks of radius $\delta\rho/2$ centered at each transmitter must be disjoint. Consider the transmitters that are within a distance between a and b from node j. The disk of radius $\delta\rho/2$ centered at each transmitter must be contained within an annulus of all points lying within a distance between $a-\frac{\delta\rho}{2}$ and $b+\frac{\delta\rho}{2}$ from receiver j. Such an annulus has an area of

$$c_3\pi\left(\left(b+\frac{\delta\rho}{2}\right)^2-\left(a-\frac{\delta\rho}{2}\right)^2\right).$$

The disks of radius $\delta\rho/2$ centered at each transmitter are disjoint, therefore the number of transmitters contained in the annulus can not be more than

$$\frac{c_3\pi((b+\frac{\delta\rho}{2})^2-(a-\frac{\delta\rho}{2})^2)}{c_4\pi(\frac{\delta\rho}{2})^2}. \tag{21}$$

In addition, the received power at j from any of the above transmitters can not exceed $\frac{P}{a^{\alpha}}$. By setting $a=k\delta\rho$, and $b=(k+1)\delta\rho$ for $k=1,2,3,...$, we can get the SIR at j is

$$\begin{aligned}
&\frac{\frac{P}{(4\rho)^{\alpha}}}{N_0+c_5\sum\limits_{k=1}^{\infty}\frac{\left((k+1)\delta\rho+\frac{\delta\rho}{2}\right)^2-\left(k\delta\rho-\frac{\delta\rho}{2}\right)^2}{\left(\frac{\delta\rho}{2}\right)^2}\frac{P}{(k\delta\rho)^{\alpha}}}\\
=&\frac{\frac{P}{4^{\alpha}}}{\rho^{\alpha}N_0+c_5\frac{P}{\delta^{\alpha}}\sum\limits_{k=1}^{\infty}\frac{16k}{k^{\alpha}}+\frac{8}{k^{\alpha}}}\\
\geq&\frac{P}{(4\rho)^{\alpha}N_0+c_5P\frac{4^{\alpha}}{\delta^{\alpha}}\left(24+\frac{16}{\alpha-2}+\frac{8}{\alpha-1}\right)}\ (\text{as }\alpha>2).
\end{aligned}$$

The above SIR is larger than β, i.e., the transmission can be successfully received, when P is sufficiently large and δ is chosen to satisfy:

$$\delta>4\left(c_5\beta\left(24+\frac{16}{\alpha-2}+\frac{8}{\alpha-1}\right)\right)^{\frac{1}{\alpha}}. \tag{22}$$

□

Lemma 5. *For a time period of T, nodes in each cell can get a total period of $\frac{T}{c}$ for transmission, and the transmission can always be successfully received by any node within the cell.*

Proof. For Voronoi cell V_i and its interfering neighbor V', there must be two points, one in V_i and one in V', that are no more than $\delta\rho$ apart. Thus, V_i and all of its interfering neighbors can be contained in a large disk of radius $(6+\delta)\rho$. This large disk can not contain more than $c_6(6+\delta)^2$ disks of radius ρ, thus V_i can not have more than $c_6(6+\delta)^2-1$ interfering neighbors.

Consider coloring all the cells such that no two interfering neighbors have the same color. A well-known fact about the vertex coloring of graphs is that a graph of degree no more than c_7 can be colored by using no more than $(1+c_7)$ colors [25]. Thus the cells can be colored by using $c_6(6+\delta)^2$ colors. We allocate a period of $\frac{T}{c_6(6+\delta)^2}$ for each color, during which the cells of that color will transmit simultaneously. As no interfering neighbors have the same color, for each period of $\frac{T}{c_6(6+\delta)^2}$, the transmission in each cell can be successfully received based on Lemma 4. □

D. Achievable Lower Bound

Based on the above results, we now present achievable lower bounds on network capacity and request satisfaction rate.

Theorem 2. *Under our network and content access model, an achievable lower bound on network capacity (averaged over all nodes) is*

$$\tilde{\lambda}(s,m,n)=\frac{sW}{m}. \tag{23}$$

Proof. Lemma 5 shows that for a time period of T, each cell can have a period of $\frac{T}{c_7}$ during which the transmission in the cell is always successfully received. Thus, for those cells that can cache all contents, during the period of $\frac{T}{c_7}$, nodes within the cell can receive $W\frac{T}{c_7}$ bits of contents regardless of which contents have been requested. Additionally, Lemma 3 has shown that at least $c_8\frac{n}{\eta}$ cells can cache all contents. Summing over all the cells that can cache all the contents, the total number of bits received is

$$c_8\frac{n}{\eta}\times W\frac{T}{c_7}=c_9\frac{WnT}{\eta}. \tag{24}$$

Then on average each node can achieve a capacity of

$$\tilde{\lambda}(s,m,n)=c_9\frac{WnT}{\eta}/nT=c_9\frac{W}{\eta}\approx\frac{sW}{m}, \tag{25}$$

as $\eta=\lceil\frac{m}{s}\rceil$. □

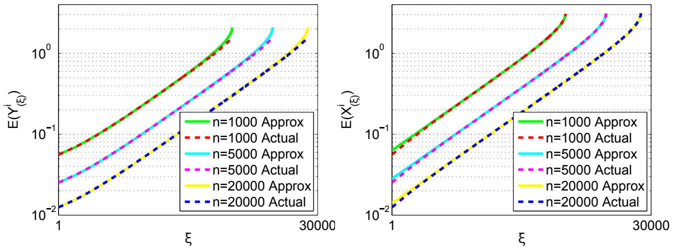
Corollary 2.1. *Under our network and content access model, an achievable lower bound on request satisfaction rate (averaged over all nodes) is*

$$\tilde{\mu}(s,m,n)=\frac{sW}{m-s}. \tag{26}$$

Proof. As each node caches s contents locally, $\frac{s}{m}$ of incoming requests will be satisfied by local cache, and the remaining requests require nodes to retrieve contents from other nodes. Theorem 2 has shown that a capacity of $\tilde{\lambda}(s,m,n)$ is achievable, which means that $\tilde{\lambda}(s,m,n)$ requests for contents not cached locally can be satisfied per second on average. An additional $\frac{s}{m-s}\tilde{\lambda}(s,m,n)$ requests for contents cached locally can be received and directly satisfied. Then,

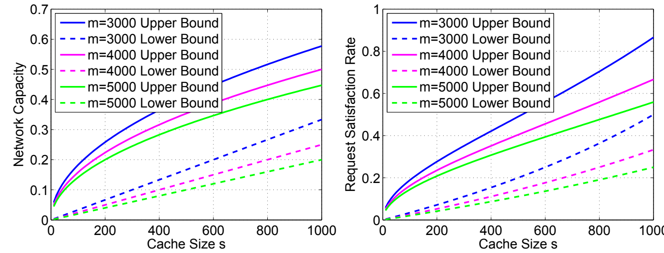
$$\tilde{\mu}(s,m,n)=\tilde{\lambda}(s,m,n)+\frac{s}{m-s}\tilde{\lambda}(s,m,n)=\frac{sW}{m-s}.$$

□



(a) $\mathrm{E}(Y^i_{(\xi)})$ (b) $\mathrm{E}(X^i_{(\xi)})$

Fig. 1: Approximation validation for $\mathrm{E}(Y^i_{(\xi)})$ and $\mathrm{E}(X^i_{(\xi)})$ (i.e., formula (7) and formula (10)).



(a) Network capacity ($n = 10^4$) (b) Request satisfaction rate ($n = 10^4$)

Fig. 2: Capacity and request satisfaction rate

VI. Numerical Results

A. Approximation Validation and Numerical Results

We first validate the two approximations we made in section IV. Fig. 1(a) verifies the approximation in formula (7), i.e., $\mathrm{E}(Y^i_{(\xi)})$ is approximated to the first term of Taylor series: $\frac{2}{3}\mathrm{E}(X^i_{(\xi)})$. We compare the actual value of $\mathrm{E}(Y^i_{(\xi)})$ with its approximated value based on the value of ξ and the number of nodes. Fig. 1(a) shows that under all three values of n, the approximation error is quite small for all possible values of ξ ($1 \leq \xi \leq n-1$). Fig. 1(b) validates the approximation in formula (10), i.e., $\mathrm{E}(X^i_{(\xi)})$ can be approximated to the first term of David and Johnson series: $\arccos(1 - \frac{2\xi}{n})$. From the figure, the difference between the actual value and the approximated value is negligible for various n and ξ.

Fig. 2 shows the effect of cache size on network capacity and request satisfaction rate based on different total sizes of unique content, where $n = 10000$. As shown in the figure, the achievable capacity lower bound grows linearly with cache size, while the capacity upper bound grows quickly when the cache size is small, and much slower when the cache size is large. We can also see that the network capacity decreases with the increase of the total size of unique content, because at this time contents have to be retrieved from faraway nodes. Fig. 2(b) shows how the cache size affects the request satisfaction rate. As can be seen, the upper bound of the request satisfaction rate grows quickly when the cache size is small, and almost linearly when the cache size is large. On the other hand, its lower bound grows linearly when the cache size is small, and grows faster when cache size is large, especially with small m ($m = 3000$). Therefore, increasing the cache size may significantly improve the request satisfaction rate when the initial cache size is either quite small or quite large. In other cases, the request satisfaction rate generally grows linearly with the cache size.

B. Comparisons to Existing Work

In this subsection, we compare our results with the analytical results presented in [10], [11], and show that our results are consistent with theirs when our caching scenario matches theirs.

(i) In case of $m = sn$, the total cache size is just enough to store all contents. To find the capacity upper bound, each content is only cached at one node. For any node i receiving a request, the requested content may be cached at any node with equal probability. Thus, if the requested content is not cached locally, it is equally likely for i to retrieve the requested content from any other node. This scenario is similar to the scenario of random network in [10], where all nodes randomly choose a destination for transmission. Our derived upper bound does conform to their results, as they present an upper bound of $O\left(\frac{W}{\sqrt{n}}\right)$ while our $\lambda(s, m, n)$ is also upper bounded by $W\sqrt{\frac{s}{m}} = \frac{W}{\sqrt{n}}$ when $m = sn$.

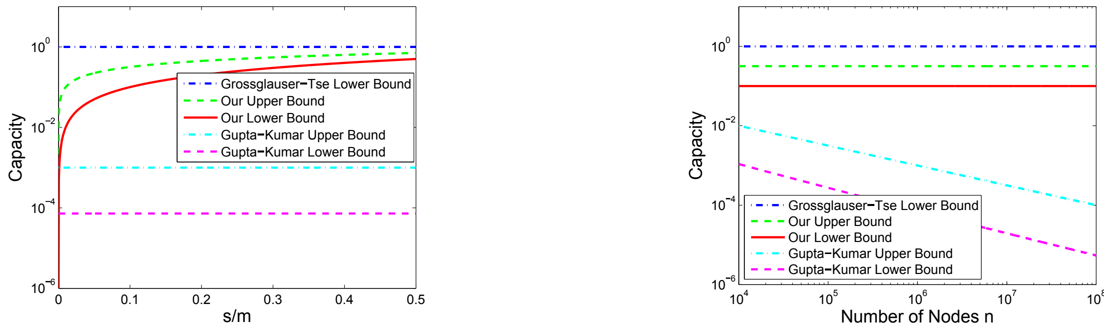
(ii) In case of $\frac{m}{2} \leq s < m$, two nodes will be enough to cache all the contents. If contents are carefully cached, each node only need to retrieve contents from its closest neighbor. This scenario is similar to the scenario constructed in [11], as communications are restricted to closest neighbors. Under this special case, we are also able to obtain an achievable capacity of $\Theta(W)$, which conforms to their results.

(iii) The most important implication of our results is that, the network capacity and the request satisfaction rate (averaged over all nodes) can remain constant even if the number of nodes increase. This is because although more nodes bring in more interferences, high node density also helps each node retrieve contents from closer neighbors.

Fig. 3 compares our results with [10], [11] based on numerical results. Fig. 3(a) shows how the capacity changes with $\frac{s}{m}$ when $n = 10^6$. Apparently, the capacity results of Gupta-Kumar [10] and Grossglauser-Tse [11] will not change with $\frac{s}{m}$. When $\frac{s}{m}$ is very small, our result is comparable to Gupta-Kumar. However, as $\frac{s}{m}$ increases, our capacity result quickly approach to Grossglauser-Tse's result. Fig. 3(b) shows the capacity as a function of the number of nodes when $\frac{s}{m} = \frac{1}{10}$. Grossglauser-Tse shows that the network capacity can remain constant through node mobility. Similarly, caching also helps achieve constant network capacity. For wireless networks without considering mobility and caching, as shown in Gupta-Kumar, the capacity decreases quickly as the number of nodes increases. Note that although Grossglauser-Tse can achieve higher network capacity than ours, they are at the cost of long delay. On the other hand, caching can increase the network capacity, and reduce the content access delay.

C. Discussions

Although our analysis has been based on the assumption that all contents have the same size of 1 bit, the proposed capacity upper and lower bounds are still valid even if contents have various sizes. For the upper bound, in the ideal scenario, each



(a) Capacity vs. cache size ($n = 10^6$)

(b) Capacity vs. the number of nodes ($\frac{s}{m} = \frac{1}{10}$)

Fig. 3: Comparisons to existing work

node still caches s bits of contents locally, and has its closest neighbors cache the remaining $m - s$ bits of contents. Thus, the average distance to retrieve contents from others ($\bar{L}$) is unchanged. Recall that the capacity upper bound is $\Theta(\frac{1}{\sqrt{n}L})$. Since $\bar{L}$ is unchanged, the capacity upper bound still holds. For the achievable lower bound, we only need to make sure that m bits can be cached in most of the Voronoi cells, which is irrelevant to the size of individual content.

Even when the nodes are distributed on a planar disk rather than the surface of the sphere as considered in this paper, we believe similar capacity upper and lower bounds can still be derived. For the capacity upper bound, the average distance to retrieve a content is still in the order of $\Theta\left(\sqrt{\frac{s}{m}}\right)$; while for the capacity lower bound, the parameters to construct Voronoi tessellation are almost the same. In addition, when the radius of the sphere increases, both the upper bound and the achievable lower bound will remain unchanged.

VII. Conclusions

In this paper, we have studied the capacity of wireless networks with caching. We proved that for nodes uniformly distributed on the surface of sphere, the network capacity is upper bounded by $\Theta(W\sqrt{\frac{s}{m}})$. We also propose a caching scheme, based on which a capacity of $\Theta(W\frac{s}{m})$ is achievable. More importantly, our results suggest that through caching, it is possible for nodes to obtain constant capacity even if the number of nodes increases.

To the best of our knowledge, this is the first paper to find an achievable capacity lower bound for wireless networks with caching. As the initial work, we do not expect to solve all the problems. In the future, we will investigate the capacity of wireless networks with caching with more practical assumptions; for example, the content request rate follows Zipf distribution.

References

[1] L. Yin and G. Cao, "Supporting cooperative caching in ad hoc networks," in *IEEE INFOCOM*, 2004.

[2] J. Zhao, P. Zhang, G. Cao, and C. R. Das, "Cooperative caching in wireless p2p networks: Design, implementation, and evaluation," *IEEE Transactions on Parallel and Distributed Systems*, vol. 21, no. 2, pp. 229–241, 2010.

[3] M. Fiore, F. Mininni, C. Casetti, and C.-F. Chiasserini, "To cache or not to cache?" in *IEEE INFOCOM*, 2009.

[4] B. Tang, H. Gupta, and S. R. Das, "Benefit-based data caching in ad hoc networks," *IEEE Transactions on Mobile Computing*, vol. 7, no. 3, pp. 289–304, 2008.

[5] E. Leonardi and G. L. Torrisi, "Least recently used caches under the shot noise mode," in *IEEE INFOCOM*, 2015.

[6] C. Fricker, P. Robert, and J. Roberts, "A versatile and accurate approximation for lru cache performance," in *International Teletraffic Congress (ITC)*, 2012.

[7] N. C. Fofack, P. Nain, G. Neglia, and D. Towsley, "Performance evaluation of hierarchical ttl-based cache networks," *Computer Networks*, vol. 65, pp. 212–231, 2014.

[8] E. J. Rosensweig, D. S. Menasche, and J. Kurose, "On the steady-state of cache networks," in *IEEE INFOCOM*, 2013.

[9] M. Dehghan, A. Seetharam, B. Jiang, T. He, T. Salonidis, J. Kurose, D. Towsley, and R. Sitaraman, "On the complexity of optimal routing and content caching in heterogeneous networks," in *IEEE INFOCOM*, 2015.

[10] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 388–404, 2000.

[11] M. Grossglauser and D. Tse, "Mobility increases the capacity of ad-hoc wireless networks," in *IEEE INFOCOM*, 2001.

[12] S. Jin and L. Wang, "Content and service replication strategies in multi-hop wireless mesh networks," in *ACM international symposium on Modeling, analysis and simulation of wireless and mobile systems (MSWiM)*, 2005.

[13] U. Niesen, D. Shah, and G. W. Wornell, "Caching in wireless networks," *IEEE Transactions on Information Theory*, vol. 58, no. 10, pp. 6524–6540, 2012.

[14] S. Gitzenis, G. Paschos, and L. Tassiulas, "Asymptotic laws for joint replication and delivery in wireless networks," *IEEE Transactions on Information Theory*, vol. 59, no. 5, pp. 2760–2776, 2013.

[15] B. Azimdoost, C. Westphal, and H. R. Sadjadpour, "On the throughput capacity of information-centric networks," in *International Teletraffic Congress (ITC)*, 2013.

[16] B. Liu, V. Firoiu, J. Kurose, M. Leung, and S. Nanda, "Capacity of cache enabled content distribution wireless ad hoc networks," in *IEEE MASS*, 2014.

[17] J. Li, C. Blake, D. S. De Couto, H. I. Lee, and R. Morris, "Capacity of ad hoc wireless networks," in *ACM MobiCom*, 2001.

[18] H. A. David and H. N. Nagaraja, *Order statistics*. Wiley Online Library, 1970.

[19] A. İ. Genç, "On order statistics from sine distribution," *Communications in StatisticsTheory and Methods*, vol. 36, no. 12, pp. 2169–2176, 2007.

[20] P. M. Burrows, "Extreme statistics from the sine distribution," *The American Statistician*, vol. 40, no. 3, pp. 216–217, 1986.

[21] A. Papoulis and S. U. Pillai, *Probability, random variables, and stochastic processes*. Tata McGraw-Hill Education, 2002.

[22] A. Okabe, B. Boots, K. Sugihara, and S. N. Chiu, *Spatial tessellations: concepts and applications of Voronoi diagrams*. John Wiley & Sons, 2009, vol. 501.

[23] H. Chernoff, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," *The Annals of Mathematical Statistics*, pp. 493–507, 1952.

[24] V. N. Vapnik and A. Y. Chervonenkis, "On the uniform convergence of relative frequencies of events to their probabilities," *Theory of Probability & Its Applications*, vol. 16, no. 2, pp. 264–280, 1971.

[25] J. A. Bondy and U. S. R. Murty, *Graph theory with applications*. Macmillan London, 1976, vol. 290.