

Multicast Capacity of Heterogeneous Networks

Qiuyu Peng, Xinbing Wang
Department of Electronic Engineering
Shanghai Jiao Tong University, China
Email: {pqy,xwang8}@sjtu.edu.cn

Abstract—In this paper, we investigate the multicast capacity for static ad hoc networks with heterogeneous clusters. We study the effect of heterogeneities on the achievable capacity from two aspects, including *heterogeneous cluster traffic* (HCT) and *heterogeneous cluster size* (HCS). HCT means cluster clients are more likely to appear near the cluster head, instead of being uniformly distributed across the network and HCS means each cluster is not equal in size as most prior literatures assume. Both of these two properties are commonly found in real networks. For this class of networks, we find that HCT increases network capacity for all the clusters and HCS does not influence the network capacity. However, smaller clusters can be allocated a much larger individual capacity than the network capacity when HCS is involved. Our work can generalize various results obtained under non-heterogeneous networks in previous literature.

Index Terms—Static Clustered Network, Multicast Capacity, Heterogeneous Cluster.

I. INTRODUCTION

A wireless network is modeled as a set of nodes that send and receive messages over a common wireless channel. Since the seminal work done by P. Gupta, P. R. Kumar [1], there is significant interest toward the asymptotic capacity of the network when the number of nodes n grows. The authors of [1] prove that the per-node capacity¹ is $\Theta(W/\sqrt{n \log n})$ in a static network. Later in [2], the capacity result is analyzed under more general fading channel condition and a similar result is given. Then M. Franceschetti, *et al.* [3] design an optimal routing protocol with capacity achieving $\Theta(W/\sqrt{n})$ via percolation theory.

The multicast network, which generalizes the above unicast network, received more attention recently and the estimation of the achievable multicast capacity is required in many applications like sensor networks and TV streaming [22]. Li, *et al.* [4] studied the achievable capacity in multicast networks. In their work, there are n multicast sessions, each comprised of 1 source and k destinations and they find that the capacity scales as $\Theta(1/\sqrt{kn \log n})$ based on the *Manhattan Routing* scheme. Their results generalize both unicast and broadcast [5] capacity. In [6], Shakkottai, *et al.* studied a different multicast framework where there are n^ϵ multicast sources and $n^{1-\epsilon}$ destinations per flow. Their network can support a rate of $\Theta(\frac{1}{\sqrt{n^\epsilon \log n}})$ for each flow. Later on, the

multicast capacity under Gaussian channel is obtained in [7], [8]. The achievable capacity in mobile multicast (motioncast) is explored in [9] and optimal mobile multicast capacity is presented in [10], which is a generalization of [11], [12]. And in [13], [14], MIMO cooperations are introduced to improve multicast capacity.

Since nodes in the same multicast session can be treated as members of a cluster, multicast networks can also be viewed as clustered networks. However, there are few works concerning the cluster behavior of multicast networks. Uniformly distributed cluster (multicast session) traffic and sizes are assumed and cluster heterogeneities are rarely involved in previous works. Actually, most real networks are characterized by various clustered heterogeneities and some aspects have already been investigated in unicast networks, which includes:

Spatial heterogeneity: Wireless nodes are not likely to be uniformly distributed across the deployed region in realistic networks e.g., wireless users may cluster in urban areas so there are less users in suburban areas. Since spontaneous grouping of the nodes around a few attraction points occurs commonly in wireless network, G. Alfano, *et al.* [15], [16] extended the capacity scaling to networks with inhomogeneous node density. In their work, nodes are generated according to a specified point process and they show that the bottleneck is in the node sparse region because the network capacity is related to the minimum node intensity.

Pattern heterogeneity: It is likely that there exists more than one type of traffic pattern in the network and nodes of the same traffic pattern constitute a cluster. In [17], Wang, *et al.* studied a unified modeling framework composed of unicast, multicast, broadcast traffic. Later in [18], Ji, *et al.* explored networks composed of both unicast and converge-cast traffic and show that MIMO cooperation can be applied to increase capacity for both types of traffic. Li, *et al.* [19] dealt with networks containing some helping nodes for packet delivery. Using this method, the normal nodes and helping nodes can be viewed as two clusters.

In retrospect, many analytical studies rely on the assumption that nodes are in i.i.d distribution. However, in many practical settings, this assumption about node distribution does not hold. Hence, the primary incentive motivates us to investigate clustered networks is because of its practical use in real life. In addition, the study of performance in clustered networks under multicast traffic is still not sufficient yet and that is another reason why we set the model. For instance, in military battlefield, commanders from different places must

¹Given two functions $f(n) > 0$ and $g(n) > 0$: $f(n) = o(g(n))$ means $\lim_{n \rightarrow \infty} f(n)/g(n) = 0$; $f(n) = O(g(n))$ means $\lim_{n \rightarrow \infty} \sup f(n)/g(n) < \infty$; $f(n) = \omega(g(n))$ is equivalent to $g(n) = o(f(n))$; $f(n) = \Omega(g(n))$ is equivalent to $g(n) = O(f(n))$; $f(n) = \Theta(g(n))$ means $f(n) = O(g(n))$ and $g(n) = O(f(n))$.

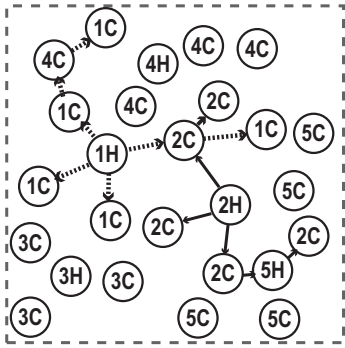


Fig. 1: Demonstration of Network topology. Nodes in the same cluster are labeled with the same number. H and D represent head (source) and clients (destination), respectively.

send requirements through a common wireless channel to their respective soldiers around them. In sensor networks, local schedulers also need to send packets to their adjacent client sensors. Since clients of the same data flow are both non-uniformly distributed and size varied, the open question is:

- What are the impacts of heterogeneous traffic and cluster size on multicast capacity in clustered networks?

In this paper, nodes in each multicast session are comprised of a cluster and the network heterogeneities include:

Heterogeneous Cluster Traffic (HCT): Clients of the same cluster (data flow) are likely to be deployed around a cluster head specified by an inhomogeneous Poisson process (IPP). We describe this clustering behavior with a variable $\sigma_{\mathcal{O}}$, which describes the extent of HCT. We find that the network capacity increases in this case because the total transmission length is shortened and we offer a quantitative relationship between the network capacity and $\sigma_{\mathcal{O}}$.

Heterogeneous Cluster Size (HCS): Clusters may have different size (cardinality) and HCS is employed to describe the population variation for each multicast data flow. We show that the network capacity is inherently related to the total number of clusters and not affected by HCS under some specific assumptions in our model.

The rest of the paper is organized as follows. In section II, we outline some preliminaries of the network and our main results. In section III and IV, a close form of the upper bound and maximized per-cluster capacity are derived respectively. In section V, we provide a routing scheme for the achievable capacity for *uniform random cluster model*. A discussion of the results is presented in section VI. Finally, we conclude this paper in section VII.

II. PRELIMINARIES AND MAIN RESULTS

A. Network Topology

We consider extended networks composed of $n_s = n^\alpha$ ($0 \leq \alpha \leq 1$) clusters distributed over a 2-dimensional torus

region \mathcal{O} of edge² $L = n^\beta$ ($0 \leq \beta \leq \alpha/2$). We specify a homogeneous Poisson process (HPP) to generate cluster head v_j , whose position is denoted by k_j for cluster C_j ($1 \leq j \leq n_s$). Then, v_j generates its cluster members according to an IPP whose intensity at ξ is given by $|\mathcal{C}_j|\phi(k_j, \xi)$, where $|\mathcal{C}_j| \leq p = n^{1-\alpha}$ is the expected size of the cluster and $\phi(\cdot)$ is the dispersion density function. In order to investigate the impact of HCT and HCS on multicast capacity, we will discuss in detail here further assumptions should be made for both $\phi(\cdot)$ and $|\mathcal{C}_j|$. Because the dispersion density function $\phi(\cdot)$ determines the client distribution of each multicast session, intuitively, HCT is mainly described by the characteristics of $\phi(\cdot)$. In addition, HCS deals with the cardinality of each cluster, so it is strongly related to $|\mathcal{C}_j|$ ($1 \leq j \leq n_s$).

First, we outline the properties of the dispersion density function $\phi(\cdot)$ as follows:

- 1) $\phi(k_j, \xi)$ is invariant under both translation and rotation with respect to k_j , therefore $\phi(k_j, \xi)$ can be rewritten as $\phi(|k_j - \xi|)$ and it is a non-increasing, non-negative, bounded, and continuous function with respect to the Euclidean distance $|k_j - \xi|$.
- 2) Integration $\phi(k_j, \xi)$ of ξ over the whole torus \mathcal{O} is equal to 1, $\int_{\mathcal{O}} \phi(k_j, \xi) d\xi = 1$.

The first property restricts the dispersion density function to a regime that clients are more likely to distribute around the cluster head, that is to say there are more clients around the cluster head and less clients in remote areas with respect to the cluster head. The second property can be derived by defining a non-negative, non-increasing continuous function $s(\rho)$ such that $\int_0^\infty \rho s(\rho) d\rho < \infty$ and then normalizing it over the whole area

$$\phi(k_j, \xi) = \frac{s(|\xi - k_j|)}{\int_{\mathcal{O}} s(|\zeta - k_j|) d\zeta}.$$

Notice that $\phi(k_j, \xi) = \Theta(s(|\xi - k_j|))$ in asymptotic analysis if we neglect the factor $\int_{\mathcal{O}} s(|\zeta - k_j|) d\zeta = \Theta(1)$.

Now we need to offer a quantitative value for a given dispersion density function $\phi(\cdot)$ to depict its degree of heterogeneity, then we can analyze the relationship between the multicast capacity and the degree of HCT. As we know, the expectation can describe the average value of a function and in this case, it corresponds to the average density distribution,

$$\mathbf{E}[\phi(x)] = \int_{\mathcal{O}} \frac{\phi(\xi)}{L^2} dx = \frac{1}{L^2}.$$

And variance can describe the fluctuation level of a function around its expectation. In this case, we define the variance of $\phi(\cdot)$ as *distribution variance* $\sigma_{\mathcal{O}}$, which can be utilized to depict HCT.

$$\sigma_{\mathcal{O}}^2 = \int_{\mathcal{O}} (\phi(\xi) - \mathbf{E}[\phi(\xi)])^2 d\xi = \int_{\mathcal{O}} \phi^2(\xi) d\xi - \frac{1}{L^2}.$$

²A cluster dense regime is assumed, which means the distance between adjacent clusters $\frac{L}{\sqrt{n_s}}$ tends to 0 when n approaches infinity. We will assume $\frac{L}{\sqrt{n_s}} = O(1)$ throughout the paper.

We omit the term k_j due to the wrap around property of a torus and it can liberate us from the border effect. In case of uniform traffic, $\phi(\xi) \equiv \frac{1}{L^2}$ therefore $\sigma_{\mathcal{O}} = 0$ and larger heterogeneity results in larger $\sigma_{\mathcal{O}}$. Finally, we specify a special point process as *uniform cluster random model* (UCRM) whose dispersion density function is as follows:

$$\phi^u(\xi) = \begin{cases} \frac{1}{\pi R^2} & |\xi| \leq R \\ 0 & \text{otherwise} \end{cases}$$

where $R = \frac{L}{\sqrt{\pi(1+(L\sigma_{\mathcal{O}})^2)}}$ is defined as cluster radius. It means clients of each cluster are randomly and uniformly distributed in a disk of radius R centered at its cluster head. We prove that this topology leads to the maximized capacity when $\sigma_{\mathcal{O}}$ is fixed in section IV.

To describe HCS, it is **relatively** simple because we can just allocate different **numbers** of nodes in each multicast session. However, we classify these n_s clusters into k super clusters (\mathcal{SC}) based on their cluster size, in order to explain the effect of cluster size. For each cluster $\mathcal{C}_j \in \mathcal{SC}_i$ ($1 \leq i \leq k$), its size $|\mathcal{C}_j| = \Theta(n^{1-\alpha_i})$, where α_i is an increasing sequence over i and the clusters in \mathcal{SC}_1 possess the largest size. In addition, some further assumptions are **shown** below.

- 1) $\alpha = \min\{\alpha_1, \alpha_2, \dots, \alpha_{k-1}, \alpha_k\} = \alpha_1$.
- 2) There are $c_0 n_s$ ($0 < c_0 < 1$) clusters in \mathcal{SC}_1 , and the other $(1 - c_0)n_s$ clusters are randomly allocated to \mathcal{SC}_i ($2 \leq i \leq k$).

The second assumption indicates that the number of clusters with size $\Theta(p)$ is the same order as the total number of clusters. Although it is somewhat restrictive here, it is constructive when we design our capacity achieving scheme and allows us to gain important insight into the structure of the problem.

Under the above assumptions, cluster clients are likely to be distributed near the cluster head and the cluster size conforms to a poisson distribution with rate $|\mathcal{C}_j|$. In addition, there are $\Theta(n_s/k)$ clusters in \mathcal{SC}_i for $2 \leq i \leq k$ applying the Chernoff bound. In order to simplify our analysis, we take both $|\mathcal{C}_j|$ and $n^{1-\alpha_i}$ as the cluster size and n_s/k as the number of clusters in \mathcal{SC}_i for $2 \leq i \leq k$. Such simplification does not influence our results in order sense. Figure 1 is an example of the network topology.

B. Transmission Protocol

All wireless transceivers can communicate over a common channel of limited bandwidth W . We adopt the protocol model for interference proposed in [1]. In each time slot, a sender i can successfully transmit at W bit/second to a destination j when the Euclidean distance between any other active transmitters and j is larger than $(1+\Delta)R_{i,j}$, where $R_{i,j}$ is the Euclidean distance between i and j ; Δ is a positive constant independent of the position of i, j, k and it specifies a guard zone for a successful transmission. Note that in broadcast cases, the interference radius is defined as $(1+\Delta)$ times the length between the furthest nodes to the source. In this paper, we assume that the minimal transmission range is denoted by

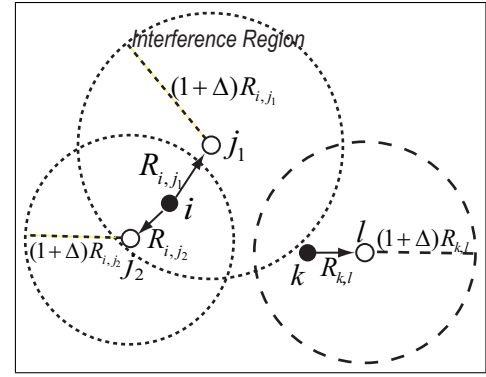


Fig. 2: Demonstration of two successful transmissions.

r and we will prove that nearest neighbor transmission is also dominant. Figure 2 illustrates two concurrent transmissions.

C. Traffic Model

A multicast traffic pattern is assumed where each cluster head generates data flows to their **respective** clients, e.g. in Figure 1. The *one to many* data flow in \mathcal{C}_j can be modeled as a multicast tree \mathcal{T}_j spanning 1 head and $|\mathcal{C}_j|$ clients. In [4], *Euclidean minimal spanning tree* (EMST) is employed to bound the length of transmission for each multicast session in non-clustered **networks** with uniform node distribution. We employ new techniques to study heterogenous clustered **networks**, which generalizes uniform cases. Let $EMST(\mathcal{C}_j)$ denote the EMST for \mathcal{C}_j and the definition is given below.

Definition of EMST: Assume that cluster head v_j generates its cluster members $\{v_j^{(1)}, v_j^{(2)}, \dots, v_j^{(|\mathcal{C}_j|)}\}$ according to an IPP. The EMST for \mathcal{C}_j in 2-dimensional space \mathbb{R}^2 connects points $v_j \cup \{v_j^{(1)}, v_j^{(2)}, \dots, v_j^{(|\mathcal{C}_j|)}\}$ using lines such that the total euclidean length of all the lines is minimized and any point can be reached from any other by following the lines.

Note that the communication between any SD pairs can also go through multiple relays from other clusters.

Since the network topology defined above is **fundamentally** related to both the number of clusters n_s and the network physical extension L , we hope to find out how the network capacity asymptotically scales. Now we give the definition of capacity.

Definition of Asymptotic Capacity: Let λ_j ($1 \leq j \leq n_s$) denote the sustainable **individual capacity** for cluster \mathcal{C}_j . Let $\lambda_s = \{\lambda_1, \lambda_2, \dots, \lambda_{n_s-1}, \lambda_{n_s}\}$ be the rate vector of all clusters. The network capacity is defined as $\lambda = \min\{\lambda_1, \lambda_2, \dots, \lambda_{n_s-1}, \lambda_{n_s}\}$. Note that the individual capacity for each cluster may differ but the network capacity is unique as defined above. Then $\lambda = \Theta(f(n))$ is defined as the asymptotic network capacity if there exist constants $c > c' > 0$, such that

$$\begin{aligned} \lim_{n \rightarrow \infty} \Pr(\lambda = cf(n) \text{ is achievable}) &< 1, \\ \lim_{n \rightarrow \infty} \Pr(\lambda = c'f(n) \text{ is achievable}) &= 1. \end{aligned}$$

Therefore λ is the minimal achievable data rate in these clusters. Let $B_j(t)$ denote the number of data units already generated in \mathcal{C}_j which have not yet been delivered to all of its members at time t . Then λ must guarantee a non-backlog network, which means $\lim_{t \rightarrow \infty} \sup_{1 \leq j \leq n_s} B_j(t) < \infty$.

D. Mathematical Notations

Throughout our paper, we denote $h_{\mathcal{C}_j}^b$ as the number of hops required for transmitting bit b to all clients in \mathcal{C}_j . ℓ_b^h is the length of transmission of bit b in its h_{th} ($1 \leq h \leq h_{\mathcal{C}_j}^b$) hop. $\delta_{h,b}$ is the number of nodes that can overhear a packet during a transmission of bit b in its h_{th} hop. $D(\xi, R)$ is the circular region centered at ξ with radius R . \mathcal{R} is the radius of the influential region centered at the head. Nodes outside the influential region cannot act as relays for that cluster. $|\mathcal{T}_j|$ is the total Euclidean length of a multicast tree \mathcal{T}_j , which can be calculated by summing up the Euclidean length of all edges belonging to the tree.

E. Useful Known Results

Throughout the paper, the following two known results would be used for proving theorems. In particular, the Chernoff bound is applied to bound the probability of sums of independent variables and the Hölder's inequality will be used for maximizing network capacity.

Lemma 2.1: Chernoff bound: Given a Poisson random variable X with expectation η , for any $\delta > 0$, the Chernoff bound is given by

$$\Pr(X > (1 + \delta)\eta) < \left(\frac{e^\delta}{(1 + \delta)^{(1 + \delta)}} \right)^\eta. \quad (1)$$

Lemma 2.2: Hölder's inequality: If S is a measurable subset of R^n with the Lebesgue measure, and f and g are measurable real- or complex-valued functions on S , then Holder inequality is

$$\int_S |f(\xi)g(\xi)| d\xi \leq \left(\int_S |f(\xi)|^p d\xi \right)^{\frac{1}{p}} \left(\int_S |g(\xi)|^q d\xi \right)^{\frac{1}{q}}, \quad (2)$$

where $1/p + 1/q = 1$.

F. Main Results

Our main results are summarized as follows:

- Given the dispersion density function $\phi(\cdot)$, the upper bound of capacity is given as follows:

$$\lambda \leq \min \left\{ W, \frac{cWL}{\sqrt{n_s} \int_{\mathcal{O}} \sqrt{\phi(\xi)} d\xi} \right\},$$

where c is some constant.

- HCT increases the maximized network capacity λ and a universal relationship between λ and $\sigma_{\mathcal{O}}$ is obtained as:

$$\lambda \leq \min \left\{ O(W), O \left(\frac{\max\{1, L\sigma_{\mathcal{O}}\}W}{\sqrt{n_s}} \right) \right\}.$$

- HCS does not affect the network capacity λ in our specified model. However, cluster \mathcal{C}_j can be allocated

$\Omega(\sqrt{p/|\mathcal{C}_j|}\lambda)$ data rate in UCRM. Therefore smaller cluster can be assigned individual capacity larger than λ and the bottleneck of increasing network capacity lies in the individual capacities of larger clusters.

III. UPPER BOUND TO MULTICAST CAPACITY

A. Derivation of Upper Bound

In this section, we will provide an upper bound of the multicast capacity, which is a crucial step for deriving the relationship between the capacity and the distribution variance. First, we will prove several key results inherent to the static network with multicast traffic. There are some tradeoffs that must be tolerated among which are the number of hops, transmission range, limited radio resources. Therefore, a thorough comprehension of the implicit relationships among them is constructive for deriving the upper bound of the achievable capacity.

It consumes radio resources to forward a bit b to relays or destinations. The following lemma captures the tradeoffs among number of hops, transmission range and limited radio resources.

Lemma 3.1: Constraint of Protocol model: Under the protocol model, the following inequality must be held for any routing scheme when the simulation time T is sufficiently large.

$$\sum_{j=1}^{n_s} \sum_{b=1}^{\lambda_j T} \sum_{h=1}^{h_{\mathcal{C}_j}^b} \frac{\pi}{16} \Delta^2 (\ell_b^h)^2 \leq WTL^2. \quad (3)$$

Proof: When T is sufficiently large, the total number of bits communicated from the head to its clients in cluster \mathcal{C}_j is $\lambda_j T$. Assume two SD pairs $X_i \rightarrow X_k$ and $X_j \rightarrow X_l$ are active in the given time slot, then according to the transmission protocol model,

$$|X_k - X_l| \geq \frac{\Delta}{2} (|X_i - X_k| + |X_j - X_l|),$$

which is derived in [1]. Thus disks of radius $\frac{\Delta}{2}$ times the transmission radius centered at the receiver can be viewed as an "exclusion region" that rejects transmitters from other data flows. Such a property also holds for broadcast that the transmission range is defined as the furthest node that can receive the packets. Let ℓ_b^h be the transmission radius centered at the receiver for the h -th hop of bit b , and S_b^h be the overlap area between the "exclusion region" of bit b 's h_{th} hop and the deployed region \mathcal{O} . Note that at least a quarter of the disk is within the given region (the extreme situation happens when the node is near the periphery of the region). Then

$$S_b^h \geq \frac{\pi}{4} \left(\frac{\Delta \ell_b^h}{2} \right)^2 = \frac{\pi \Delta^2 (\ell_b^h)^2}{16}.$$

Therefore, radio resources can be viewed as the limited bandwidth W times the simulation time T and the network area L^2 , such that the following inequality is satisfied:

$$\sum_{j=1}^{n_s} \sum_{b=1}^{\lambda_j T} \sum_{h=1}^{h_{\mathcal{C}_j}^b} \frac{\pi}{16} \Delta^2 (\ell_b^h)^2 \leq \sum_{j=1}^{n_s} \sum_{b=1}^{\lambda_j T} \sum_{h=1}^{h_{\mathcal{C}_j}^b} S_b^h \leq WTL^2.$$

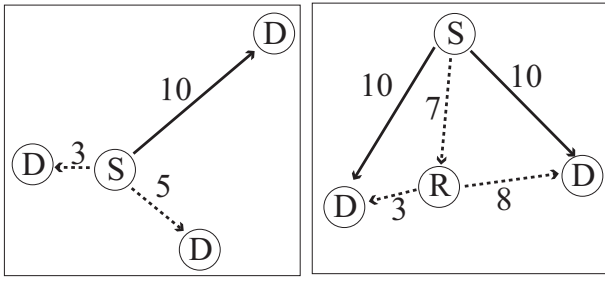


Fig. 3: Demonstration of the two conditions for MTTL.

The above inequality is a basic requirement for multihop transmission type and the cooperative MIMO as in [14] is not considered here. For simplicity, we define the area of the exclusion region consumed by one bit in a certain hop as a single unit of transmission resources. The left side sums up all the transmission resources consumed by each hop of each bit b in each cluster \mathcal{C}_j . Since at most WT bits can be carried in time T with bandwidth W , for area L^2 , the upper bound of total transmission resource is $WT \cdot L^2 = WTL^2$. This inequality cannot only be used in our heterogeneous case, but also in homogeneous case as well since it is derived regardless of node distribution. In static network, there is a minimal transmission range r to guarantee full network connectivity. In networks with uniform node distribution, it is proved that the transmission range $r = \Omega(\sqrt{\log n/n})$ is sufficient for network connectivity *w.h.p.* in [1]. Now we need to characterize a feasible transmission range r , below which the transmission is not possible in our framework. In other words, we want to derive r such that the number of nodes in $D(\xi, r)$ is nonzero all the time.

Lemma 3.2: Let $\mathcal{N}(r)$ represent the number of nodes within the transmission range r when a node wants to transmit its packets, then $\mathcal{N}(r) = \Theta(\frac{n_s p r^2}{L^2})$ when $r = \Omega(\frac{L}{\sqrt{n_s p}})$.

Proof: In our cluster dense regime, when $r = \omega(\frac{1}{\sqrt{n_s p/L}})$, $\mathcal{N}(r) \leq \frac{2\pi \sum_{j=1}^{n_s} |\mathcal{C}_j| r^2}{L^2} \leq \frac{2\pi n_s p r^2}{L^2}$ based on Chernoff bound and Riemann sum (one can refer to Theorem 1 in [15]). Similarly, the lower bound of $\mathcal{N}(r)$ is as $\mathcal{N}(r) \geq \frac{c_0 \pi n_s p r^2}{2L^2}$.

Then we come to $r = \Theta(\frac{L}{\sqrt{n_s p}})$ case. Note that the node density is upper bounded by a HPP with rate $\mu = \frac{\pi r^2 n_s p}{L^2}$, the probability that the number of nodes inside the disk exceeding n_0 is

$$\Pr(\mathcal{N}(r) > n_0) \leq e^{-\mu} \sum_{i=n_0+1}^{\infty} \frac{\mu^i}{i!} \leq \frac{e^{(\theta-1)\mu} \mu^{n_0+1}}{(n_0+1)!}.$$

During the above derivation, we used the Lagrange form of the remainder term and $0 \leq \theta \leq 1$. Therefore when $n_0 = \omega(\frac{n_s p r^2}{L^2})$, $\Pr(\mathcal{N}(r) > n_0) = 0$ *w.h.p.* Note that $\mathcal{N}(r) \geq 1$ is a prerequisite for any transmissions and we complete our proof. ■

Therefore we know a necessary condition for the transmission range r is $\ell_b^h \geq r = \Omega(\frac{L}{\sqrt{n_s p}})$. Now we need to

characterize the total length for transmitting a bit $\sum_{h=1}^{h_b} \ell_b^h$. In [4], they prove that nearest neighbor transmission can achieve optimal capacity in uniform traffic distribution. They obtain that the number of hops required is $\sqrt{k\bar{n}}$ for k destinations. The situation is complicated here since traffic is not uniformly distributed across the network. Clients in the same multicast session are more clustered around the head and a larger transmission range can cover more clients in a cluster when the transmission happens near the cluster head (source) than other places. Therefore, it is unknown whether the many more destinations involved in a larger transmission range can compensate for the sacrifice of radio resources. Unfortunately, the following analysis reveals that nearest neighbor transmission is also optimal in HCT.

In [4], EMST is investigated in multicast traffic and it can help us obtain the capacity upper bound in non-heterogeneous networks. To obtain the results under heterogeneous cluster traffic and size, we first introduce the following lemma. Theorem 1 in [21] as follows.

Lemma 3.3: If f is the density of the probability function for picking points, then for large n and $d \neq 1$, the size of the EMST is approximately $c(d)n^{\frac{d-1}{d}} \int_{\mathbb{R}^d} f(x)^{\frac{d-1}{d}} dx$, where $c(d)$ is a constant depending only on the dimension d .

Our case corresponds to $d = 2$ and such that

$$|EMST(\mathcal{C}_j)| = \Theta\left(\sqrt{|\mathcal{C}_j|} \int_{\mathcal{O}} \sqrt{\phi(\xi)} d\xi\right).$$

It reveals that the length of the Euclidean minimum multicast

Algorithm 1 Generation of ρ -simplified EMST from EMST.
Input: $\mathcal{C}_j, \phi(\cdot)$ Output: $\rho - \mathcal{C}_j$

- 1: Specify two point sets $S \leftarrow \emptyset, S' \leftarrow \mathcal{C}_j$.
- 2: Randomly label each node in S' with numbers $1, 2, \dots, |\mathcal{C}_j| - 1, |\mathcal{C}_j|$.
- 3: Choose nodes with the smallest label number in S' .
- 4: Add the chosen nodes to S and discard all the nodes within $D(\xi', \rho)$ in S' , where ξ' is the position of the chosen node.
- 5: Back to step 3 until no node is left in S' , $(\rho - \mathcal{C}_j) \leftarrow S'$.

spanning tree constitutes two terms, the square root of the number of nodes $\sqrt{|\mathcal{C}_j|}$ and $\int_{\mathcal{O}} \sqrt{\phi(\xi)} d\xi$. Note that we eliminate the constants $c(d)$ for simplicity. Intuitively, the minimum length for connecting all the nodes is the minimum total transmission length (MTTL). For a given graph, MTTL can be counted by summing all the irredundant transmission lengths with extra intermediate nodes and edges. Although MTTL is superficially similar to the EMST problem, the length of MTTL is far less than EMST for the following reasons.

- Larger transmission range can cover more than one node, but only the length of the longest SD pair is counted. For instance, if a node broadcast its message to all the other nodes in one time, the MTTL is at most L . Therefore MTTL is related to the transmission range r .
- Nodes from other clusters can act as relays to help forward information. We must consider the impact of

relays on the MTTL.

Figure 3 illustrates two examples of the above two questions, respectively. In order to answer the first question (derive the relationship between the transmission range r and MTTL), we construct a new concept ρ -simplified cluster $\rho - \mathcal{C}_j$. It is used to generate a less dense node distribution given \mathcal{C}_j and $\phi(\cdot)$. Algorithm 1 illustrates how to generate $\rho - \mathcal{C}_j$ given \mathcal{C}_j and the dispersion density function $\phi(\cdot)$.

Let $\rho - \mathcal{T}_j$ denote the multicast tree spanning $\rho - \mathcal{C}_j$. The length of each branch in $\rho - \mathcal{T}_j$ is larger than ρ and all the abandoned nodes are within a distance ρ from the nodes in $\rho - \mathcal{C}_j$. $|\rho - \mathcal{T}_j| \geq |EMST(\rho - \mathcal{C}_j)|$ according to the definition of EMST. After the thinning process according to lemma 3.3, we specify two regions according to $\phi(\cdot)$. Let $\phi'(\xi)$ denote the point intensity after Algorithm 1.

- **Dense Region** ($S_{1,j}$) Nodes in this region are populous and we specify a radius

$$\tilde{\rho}_j = \sup\{\rho_j, \phi(\rho_j) \geq \frac{1}{\pi\rho^2|\mathcal{C}_j}|\}$$

for this circular region $S_{1,j}$ because $\phi(\cdot)$ is invariant under rotations. After the thinning process, $\phi'(\xi) \geq \Theta(\frac{1}{\pi\rho^2|\mathcal{C}_j})$ on the basis of Chernoff bound.

- **Sparse Region** ($S_{2,j} = \mathcal{O}/S_{1,j}$) Nodes in this region are relatively sparse such that there are at most a constant number of nodes in $D(\xi, \rho)$ if $\xi \in S_{2,j}$. After the thinning process, the node density is at least a constant fraction of the original density. Therefore $\phi'(\xi) \geq \Theta(\phi(\xi))$.

Lemma 3.4: There exists a constant $c_2 > 0$ such that $|EMST(\rho - \mathcal{C}_j)|$ can be lower bounded as

$$|EMST(\rho - \mathcal{C}_j)| \geq c_2\sqrt{|\mathcal{C}_j|} \left(\frac{\sqrt{\pi}\tilde{\rho}_j^2}{\rho} + \int_{S_{2,j}} \sqrt{\phi'(\xi)}d\xi \right).$$

Proof: The length of EMST is determined by the point intensity according to lemma 3.3, thus there exists a constant $c' > 0$, such that

$$\begin{aligned} |EMST(\rho - \mathcal{C}_j)| &\geq c' \int_{\mathcal{O}} \sqrt{|\mathcal{C}_j|\phi'(\xi)}d\xi \\ &= c' \sqrt{|\mathcal{C}_j|} \left(\int_{S_{1,j}} \sqrt{\phi'(\xi)}d\xi + \int_{S_{2,j}} \sqrt{\phi'(\xi)}d\xi \right) \\ &\geq c_2 \sqrt{|\mathcal{C}_j|} \left(\frac{\sqrt{\pi}\tilde{\rho}_j^2}{\rho\sqrt{|\mathcal{C}_j|}} + \int_{S_{2,j}} \sqrt{\phi(\xi)}d\xi \right), \end{aligned}$$

where c_2 is also a constant. Note that our result holds even when $S_{1,j}$ or $S_{2,j}$ is empty. ■

Now we come to the second question. The following lemma tells us that $|EMST(\rho - \mathcal{C}_j)|$ is at most $\frac{\sqrt{3}}{2}$ times larger than MTTL when relays are utilized, which is proved in [20].

Lemma 3.5: Given k nodes U , any multicast tree spanning these k nodes (may be using some additional relay nodes) will have an Euclidean length of at least $\varrho|EMST(U)|$, where $\varrho = \sqrt{3}/2$ and $|EMST(U)|$ is the EMST spanning U .

Then we obtain the lower bound of MTTL for \mathcal{C}_j as

$$MTTL(\mathcal{C}_j) \geq \Theta(|EMST(r - \mathcal{C}_j)|).$$

Based on the above inequality, we obtain a lower bound of $h_{\mathcal{C}_j}^b$ as:

$$h_{\mathcal{C}_j}^b \geq \frac{MTTL(\mathcal{C}_j)}{r} \geq \frac{\Theta(|EMST(r - \mathcal{C}_j)|)}{r} \quad (4)$$

Note that what we are interested in is the order sense of the result, therefore we directly regard r as the transmission length. Now we will investigate the upper bound of network capacity λ based on the previous analysis of the restrictions imposed by the network. The results obtained here are some fundamental limits that cannot be violated by any routing protocols.

Theorem 3.1: Under the assumptions of the proposed wireless network, the following tradeoffs must be satisfied by all scheduling policies.

$$\sum_{j=1}^{n_s} \lambda_j \sqrt{|\mathcal{C}_j|} \leq c \frac{\sqrt{n_s p W L}}{\int_{\mathcal{O}} \sqrt{\phi(\xi)}d\xi}, \quad (5)$$

where $\lambda_j = O(W)$ and c is a constant.

Proof: Using the Cauchy-Schwartz inequality and Eqn.(3), we have

$$\begin{aligned} \left(\sum_{j=1}^{n_s} \sum_{b=1}^{\lambda_j T} \sum_{h=1}^{h_{\mathcal{C}_j}^b} \ell_b^h \right)^2 &\leq \left(\sum_{j=1}^{n_s} \sum_{b=1}^{\lambda_j T} \sum_{h=1}^{h_{\mathcal{C}_j}^b} (\ell_b^h)^2 \right) \left(\sum_{j=1}^{n_s} \sum_{b=1}^{\lambda_j T} \sum_{h=1}^{h_{\mathcal{C}_j}^b} 1 \right) \\ &\leq \frac{16WTL^2}{\pi\Delta^2} \left(\sum_{j=1}^{n_s} \sum_{b=1}^{\lambda_j T} h_{\mathcal{C}_j}^b \right) \quad (6) \end{aligned}$$

Because r is the minimal transmission range in the network, the following inequality is satisfied.

$$h_{\mathcal{C}_j}^b \leq \frac{\sum_{h=1}^{h_{\mathcal{C}_j}^b} \ell_b^h}{r},$$

and substitute it into inequality (6), we can obtain that

$$\sum_{j=1}^{n_s} \sum_{b=1}^{\lambda_j T} h_{\mathcal{C}_j}^b \leq \frac{16WTL^2}{\pi\Delta^2 r^2}. \quad (7)$$

On the other hand, we will get a lower bound on the left side of the above inequality. First we divide n_s clusters into two sets $\mathbb{S}_1, \mathbb{S}_2$ as:

$$\begin{aligned} \mathbb{S}_1 &= \{\mathcal{C}_j | \text{Dense Region } S_{1,j} \neq \emptyset\}, \\ \mathbb{S}_2 &= \{\mathcal{C}_j | \text{Dense Region } S_{1,j} = \emptyset\}. \end{aligned}$$

Based on eqn. (4) and the fact that $\frac{|EMST(\mathcal{C}_j)|}{\sqrt{|\mathcal{C}_j|}} = \Theta\left(\int_{\mathcal{O}} \sqrt{\phi(\xi)}d\xi\right)$ given in Lemma 3.3, we obtain that if

$\mathbb{S}_1 \neq \emptyset$, there exists a constant $c_4, c'_4, c_5 > 0$, such that

$$\begin{aligned}
\sum_{j=1}^{n_s} \sum_{b=1}^{\lambda_j T} h_{\mathcal{C}_j}^b &\geq \frac{c_4}{r} \sum_{j=1}^{n_s} \sum_{b=1}^{\lambda_j T} |EMST(r - \mathcal{C}_j)| \\
&\geq \frac{c_4}{r} \left(\sum_{\mathcal{C}_j \in \mathbb{S}_1} \sum_{b=1}^{\lambda_j T} |EMST(r - \mathcal{C}_j)| \right. \\
&\quad \left. + \sum_{\mathcal{C}_j \in \mathbb{S}_2} \sum_{b=1}^{\lambda_j T} |EMST(r - \mathcal{C}_j)| \right) \\
&\geq \frac{c_4 T}{r} \left(\sum_{\mathcal{C}_j \in \mathbb{S}_1} \lambda_j \sqrt{|\mathcal{C}_j|} \right) \psi(\phi(\cdot), \mathbb{S}_1) \\
&\quad + \frac{c'_4 T}{r} \left(\sum_{\mathcal{C}_j \in \mathbb{S}_2} \lambda_j \sqrt{|\mathcal{C}_j|} \right) \psi(\phi(\cdot), \mathbb{S}_1) \\
&\geq \frac{c_5 T}{r} \left(\sum_{j=1}^{n_s} \lambda_j \sqrt{|\mathcal{C}_j|} \right) \psi(\phi(\cdot), \mathbb{S}_1), \quad (8)
\end{aligned}$$

where

$$\psi(\phi(\cdot), \mathbb{S}_1) = \min_{\mathcal{C}_j \in \mathbb{S}_1} \left\{ \frac{|EMST(r - \mathcal{C}_j)|}{\sqrt{|\mathcal{C}_j|}} \right\}.$$

Then by substituting Eqn. (7) into Eqn. (8), we can obtain

$$\begin{aligned}
\sum_{j=1}^{n_s} \lambda_j \sqrt{|\mathcal{C}_j|} &\leq \frac{r}{c_5 T \psi(\phi(\cdot), \mathbb{S}_1)} \sum_{j=1}^{n_s} \sum_{b=1}^{\lambda_j T} h_{\mathcal{C}_j}^b \\
&\leq \frac{4\sqrt{2}WL^2}{\Delta\sqrt{c_0}c_5 r \psi(\phi(\cdot), \mathbb{S}_1)}. \quad (9)
\end{aligned}$$

According to the definition of capacity,

$$\begin{aligned}
\sum_{j=1}^{n_s} \lambda_j \sqrt{|\mathcal{C}_j|} &\geq \lambda \sum_{j=1}^{n_s} \sqrt{|\mathcal{C}_j|} \\
&= \lambda \left(\sum_{\mathcal{C}_j \in \mathcal{SC}_1} \sqrt{|\mathcal{C}_j|} + \sum_{\mathcal{C}_j \notin \mathcal{SC}_1} \sqrt{|\mathcal{C}_j|} \right) \quad (10)
\end{aligned}$$

Recalling the previous assumption that the number of clusters in \mathcal{SC}_1 with size $\Theta(p)$ is the same order with total cluster numbers n_s , we have $\sum_{\mathcal{C}_j \in \mathcal{SC}_1} \sqrt{|\mathcal{C}_j|} = \Theta(n_s \sqrt{p})$. Hence, there exists a constant c_6 such that

$$\begin{aligned}
\lambda \left(\sum_{\mathcal{C}_j \in \mathcal{SC}_1} \sqrt{|\mathcal{C}_j|} + \sum_{\mathcal{C}_j \notin \mathcal{SC}_1} \sqrt{|\mathcal{C}_j|} \right) &\geq c_6 \lambda n_s \sqrt{p} \\
&\quad + \lambda \sqrt{n^{1-\alpha_k}} (1 - c_0) n_s \\
&\geq c_6 \lambda n_s \sqrt{p}. \quad (11)
\end{aligned}$$

Combine Eqn. 9 and Eqn. 10, we obtain

$$\lambda \leq \frac{4\sqrt{2}WL^2}{\Delta\sqrt{c_0}c_5 c_6 n_s \sqrt{p} r \psi(\phi(\cdot), \mathbb{S}_1)}.$$

Else if $\mathbb{S}_1 = \emptyset$:

$$\begin{aligned}
\sum_{j=1}^{n_s} \sum_{b=1}^{\lambda_j T} h_{\mathcal{C}_j}^b &\geq \frac{c_4}{r} \sum_{j=1}^{n_s} \sum_{b=1}^{\lambda_j T} |EMST(r - \mathcal{C}_j)| \\
&\geq \frac{c_4}{r} \sum_{j=1}^{n_s} \sum_{b=1}^{\lambda_j T} \sqrt{|\mathcal{C}_j|} \int_{\mathcal{O}} \sqrt{\phi(\xi)} d\xi \\
&= \frac{c_4 T}{r} \int_{\mathcal{O}} \sqrt{\phi(\xi)} d\xi \left(\sum_{j=1}^{n_s} \lambda_j \sqrt{|\mathcal{C}_j|} \right). \quad (12)
\end{aligned}$$

Substitute Eqn. (7) and Eqn.(11) into (12), we obtain

$$\begin{aligned}
c_6 \lambda n_s \sqrt{p} &\leq \frac{r}{c_4 T \int_{\mathcal{O}} \sqrt{\phi(\xi)} d\xi} \sum_{j=1}^{n_s} \sum_{b=1}^{\lambda_j T} h_{\mathcal{C}_j}^b \\
&\leq \frac{4\sqrt{2}WL^2}{\Delta\sqrt{c_0}c_4 r \int_{\mathcal{O}} \sqrt{\phi(\xi)} d\xi} \\
&\leq \frac{4WL\sqrt{2n_s p}}{\Delta c' \sqrt{c_0} c_4 \int_{\mathcal{O}} \sqrt{\phi(\xi)} d\xi}. \quad (13)
\end{aligned}$$

During the above derivation, $r \geq c' \frac{L}{\sqrt{n_s p}}$ is a necessary condition to guarantee network connectivity, where $c' > 0$ is a constant. After comparing the results in the two cases, the only thing required to do is to prove that there exists a constant $c_5 > 0$ such that

$$\frac{L\sqrt{n_s p}}{c' \int_{\mathcal{O}} \sqrt{\phi(\xi)} d\xi} \geq \frac{c_7 L^2}{r \psi(\phi(\cdot), \mathbb{S}_1)}.$$

Recalling Lemma 3.4, it is equivalent to prove

$$c' c_7 \int_{\mathcal{O}} \sqrt{\frac{\phi(\xi)}{n_s p / L^2}} d\xi \leq c_2 \tilde{p}_j^2 + r \int_{S_{2,j}} \sqrt{\phi(\xi)} d\xi.$$

Note that $r \geq c' L / \sqrt{n_s p}$ and $\phi(\xi) \leq \frac{2p}{pL^2} = \Theta(1/L^2)$ (See Section III.B), and there exists a constant c_7 that can meet the above inequality and c satisfies Eqn. (5). ■

Theorem 3.2: The achievable capacity λ in our network is upper bounded as follows:

$$\lambda \leq \min \left\{ W, \frac{cWL}{\sqrt{n_s} \int_{\mathcal{O}} \sqrt{\phi(\xi)} d\xi} \right\}. \quad (14)$$

Proof: Since we assume that the number of clusters with size $\Theta(p)$ is in the same order with total number of clusters n_s . we have

$$\sum_{j=1}^{n_s} \lambda_j \sqrt{|\mathcal{C}_j|} \geq \lambda \sum_{j=1}^{n_s} \sqrt{|\mathcal{C}_j|} = \Theta(\lambda n_s \sqrt{p}).$$

Combining with Theorem 3.1, the term \sqrt{p} can be eliminated. Hence, We have also proved why HCS has no influence on the network capacity in our specified model. ■

B. Discussion on Heterogeneous Traffic Size

From previous results, we know that heterogeneous traffic size does not affect network capacity under our specified conditions. Furthermore, we find that, although the node distribution in each cluster conforms to an inhomogeneous **Poisson** process, the node distribution of the whole network seems to conform to a homogenous **Poisson** process. We first introduce a lemma which will be used in our proof later.

Lemma 3.6: Consider a set of n_s points distributed over a 2-dimensional domain \mathcal{O} of area L^2 according to an HPP. Let $\mathcal{A} = \{A_i\}$ be a regular square tessellation of \mathcal{O} , whose elements have a area $|A_i|$. Let $U(A_i)$ be the number of points falling within A_i . Then $\frac{n_s|A_i|}{2L^2} < \inf_k U(A_i) < \sup_k U(A_i) < \frac{2n_s|A_i|}{L^2}$ w.h.p.

Proof: The proof follows from Chernoff's bound. Let X be the number of nodes in squarelet A_i . By substituting $\delta = 1$ and $\eta = \frac{n_s|A_i|}{L^2}$ in to Eqn. (1), we have

$$\Pr(X > \frac{2n_s|A_i|}{L^2}) < \left(\frac{e}{4}\right)^{\frac{n_s|A_i|}{L^2}} \rightarrow 0$$

as n_s approaches infinity. Similarly, the left side can be proved by applying Chernoff bound. ■

Theorem 3.3: Let $\underline{\phi}$ and $\bar{\phi}$ denote the minimum and maximum node density across the network area \mathcal{O} . p and \underline{p} represent the largest and smallest cluster size in \mathcal{SC}_1 . Then

$$\frac{p}{2pL^2} \leq \underline{\phi} \leq \bar{\phi} \leq \frac{2p}{\underline{p}L^2},$$

when $\frac{\sqrt{n_s}}{L} = \Omega(1)$. Note that p and \underline{p} are of the same order under our constraints.

Here we present detailed proof as follows.

Proof: Consider a generic point ξ_0 . The node density at ξ_0 is denoted by $\Phi(\xi_0)$ and can be expressed by the sum of each cluster's contribution to ξ_0 .

$$\Phi(\xi_0) = \sum_{j=1}^{n_s} |C_j| \phi(k_j, \xi_0) = \sum_{j=1}^{n_s} |C_j| \frac{s(|\xi_0 - k_j|)}{\int_{\mathcal{O}} s(|\zeta - k_j|) d\zeta}.$$

We also define another density $\Phi'(\xi_0) = \sum_{C_j \in \mathcal{SC}_1} |C_j| \phi(k_j, \xi_0)$ in terms of clusters allocated in \mathcal{SC}_1 (i.e., those with cluster size $\Theta(p)$). Obviously,

$$\Phi'(\xi_0) \leq \Phi(\xi_0).$$

Let \underline{d}_{0i} and \bar{d}_{0i} be the inferior and superior distances between points $\xi \in A_i$ and ξ_0 , and $\underline{U}(A_i)$ and $\bar{U}(A_i)$ be a lower bound and an upper bound to the number of cluster heads falling in A_i , respectively. Then

$$\sum_i \underline{p} \frac{s(\bar{d}_{0i})}{H} \underline{U}(A_i) \leq \Phi'(\xi_0) \leq \Phi(\xi_0) \leq \sum_i p \frac{s(\underline{d}_{0i})}{H} \bar{U}(A_i),$$

being $H = \int_{\mathcal{O}} s(|\zeta - k_j|) d\zeta$.

Applying the above lemma 3.6, we have that

$$\frac{n_s}{2L^2} \underline{p} \sum_i \frac{s(\bar{d}_{0i})}{H} |A_i| \leq \Phi'(\xi_0) \leq \Phi(\xi_0) \leq \frac{2n_s}{L^2} p \sum_i \frac{s(\underline{d}_{0i})}{H} |A_i|.$$

Note that when the area of squarelet vanishes to 0,

$$\begin{aligned} \sum_i \frac{s(\bar{d}_{0i})}{\int_{\mathcal{O}} s(|\zeta - k_j|) d\zeta} |A_i| &\sim \sum_i \frac{s(\underline{d}_{0i})}{\int_{\mathcal{O}} s(|\zeta - k_j|) d\zeta} |A_i| \\ &\sim \int_{\mathcal{O}} s(|\zeta - \xi_0|) d\zeta = 1. \end{aligned}$$

Hence, the node density at ξ_0 can be bounded by

$$\frac{n_s}{2L^2} \underline{p} \leq \Phi(\xi_0) \leq \frac{2n_s}{L^2} p.$$

Consequently, we obtain the result by dividing both sides of the above inequality by $n_s p$ and $n_s \underline{p}$, respectively. ■

IV. MAXIMIZED CAPACITY WITH DISTRIBUTION VARIANCE CONSTRAINED

In the previous section, we have completed our analysis of the relationship between the network capacity λ and the dispersion density function $\phi(\cdot)$. However, we cannot tell what the impact is on the achievable capacity. Therefore a quantitative value $\sigma_{\mathcal{O}}$ is needed to describe the degree of heterogeneity of each dispersion density function $\phi(\cdot)$. There are two reasons for studying their relationship. First, we cannot see whether heterogeneity increases or decreases network capacity λ based on Eqn. (14). $\sigma_{\mathcal{O}}$ is a good indicator of the degree of heterogeneity. Second, we can predict the maximized achievable capacity just by knowing the distribution variance. The exact point process is not needed using this model. This is useful because it is often difficult to obtain the dispersion density function.

Recalling the definition of $\sigma_{\mathcal{O}}$, there are various dispersion density functions $\phi(\cdot)$ satisfied given a fixed $\sigma_{\mathcal{O}}$. Below we will show how to find the point process that maximizes network capacity.

Theorem 4.1: Given the distribution variance $\sigma_{\mathcal{O}}$, the maximized network capacity λ is bounded as follows:

$$\lambda \leq \min \left\{ O(W), O \left(\frac{\max\{1, L\sigma_{\mathcal{O}}\} W}{\sqrt{n_s}} \right) \right\}. \quad (15)$$

The respect dispersion function is identical to $\phi^u(\xi)$.

According to Theorem 3.1, a smaller $|EMST(C_j)|$ results in a larger capacity. Therefore we will derive the minimum $|EMST(C_j)|$ given a fixed $\sigma_{\mathcal{O}}$.

Theorem 4.2: Define a real variable function $\Upsilon(\phi(\cdot)) = \int_{\mathcal{O}} \sqrt{\phi(\xi)} d\xi$. Then we can prove that $\phi^u(\cdot)$ can minimize $\Upsilon(\phi(\cdot))$ for all $\phi(\cdot)$ with distribution variance $\sigma_{\mathcal{O}}$.

Proof: We will refer to Hölder's inequality for the proof of this theorem as follows.

Let $\sigma_{\mathcal{O}}^2 = \frac{1}{\pi R^2} - \frac{1}{L^2}$ and we can obtain that

$$\int_{\mathcal{O}} \phi^2(\xi) d\xi \leq \frac{1}{\pi R^2}.$$

Therefore our minimization problem is transformed as follows:

$$\begin{aligned} \min \quad & \Upsilon(\phi(\cdot)) = \int_{\mathcal{O}} \sqrt{\phi(\xi)} d\xi \\ \text{s.t.} \quad & \int_{\mathcal{O}} \phi(\xi) d\xi = 1 \\ & \int_{\mathcal{O}} \phi^2(\xi) d\xi \leq \frac{1}{\pi R^2} \\ & \phi(|\xi_1|) \leq \phi(|\xi_2|) \text{ for all } |\xi_1| \geq |\xi_2| \end{aligned}$$

We substitute $f(\xi) = \phi^{2/3}(\xi)$, $g(\xi) = \phi^{1/3}(\xi)$, $p = 3$ and $q = 3/2$ into Eqn. (2). Then we can obtain the following inequality in space \mathcal{O} .

$$\begin{aligned} \int_{\mathcal{O}} \sqrt{\phi(\xi)} d\xi & \geq \left(\frac{\int_{\mathcal{O}} \phi(\xi) d\xi}{\left(\int_{\mathcal{O}} \phi^2(\xi) d\xi\right)^{1/3}} \right)^{3/2} \\ & \geq \sqrt{\pi} R. \end{aligned}$$

Note that $\phi(\cdot)$ is a monotonously decreasing function and the inequality will become an equation when

$$\phi(\xi) = \begin{cases} \frac{1}{\pi R^2} & |\xi| \leq R \\ 0 & \text{otherwise} \end{cases}$$

which means that $\phi^u(\cdot)$ can achieve the minimal value of $\Upsilon(\phi(\cdot))$. ■

Recalling Theorem 3.2, we complete the proof of Theorem 4.1. In this case, the node distribution conforms **exactly** to the proposed UCRM. In this model, cluster members are uniformly and randomly distributed in a disk of radius R centered at the cluster head. In **the** next section, we will provide the capacity achieving scheme to approach this bound and verify the maximized capacity is achievable in order sense.

V. CAPACITY ACHIEVING SCHEME FOR UNIFORM CLUSTER RANDOM MODEL

We have already derived maximized multicast capacity upper bound when two types of heterogeneities are involved and we proved that UCRM is the node distribution that can achieve the maximized capacity. In this section, we will provide a multicast capacity achieving scheme for UCRM based on percolation theory. We prove that the multicast capacity achieved by our scheme matches the asymptotic upper bounds when the number of sessions n_s is large enough.

We find that only when the length of the multicast s-spanning tree $|\mathcal{T}_j|$ is on the same order with $|EMST(\mathcal{C}_j)|$, the per-cluster capacity can approach the theoretical upper bound in order sense. Intuitively, when the degree of HCT is high enough, the nodes in each cluster are only overlapped with a constant number of clusters, which corresponds to $\sigma_{\mathcal{O}} = \Omega(\frac{\sqrt{n_s}}{L})$. We denote such a case as **trivial cluster overlapping**. On the contrary, when the degree of HCT is **relatively** low, all the nodes are fully overlapped and we find that the achievable capacity is identical to the uniform case when $\sigma_{\mathcal{O}} = O(\frac{1}{L})$. We denote such a case as **fully cluster**

overlapping. The network topology between these two extreme cases is denoted by **partial cluster overlapping**. The capacity achieving scheme for uniform cases has been developed in [4] and we discuss the other two cases in this section.

A. When $\sigma_{\mathcal{O}} = \Omega(\frac{\sqrt{n_s}}{L})$

In this case, $R = \frac{L}{\sqrt{\pi(1+L^2(\sigma_{\mathcal{O}})^2)}} = O(\frac{L}{\sqrt{n_s}})$. There are at most a constant number of clusters inside $D(\xi, R)$ for $\xi \in \mathcal{O}$ and a simple TDMA scheme can achieve $\Theta(W)$ capacity for each cluster.

B. When $\sigma_{\mathcal{O}} = o(\frac{\sqrt{n_s}}{L})$

In this case, $R = \Theta(\frac{1}{\sigma_{\mathcal{O}}}) = \omega(\frac{L}{\sqrt{n_s}})$ and the traffic in each cluster is not so aggregated because $\sigma_{\mathcal{O}}$ is **relatively** smaller. In [3], an **information highway** is proposed to approach the capacity upper bound for unicast non-clustered **networks** based on percolation theory and we apply this concept to our routing scheme. Before we outline **the definition of an information highway**, some useful lemmas are provided.

Lemma 5.1: Let $\mathcal{N}(r)$ denote the number of nodes within a disk of radius $r = \Theta(L/\sqrt{n_s p})$. Then if the cluster radius $R = \Omega(L/\sqrt{n_s})$, the following inequality is satisfied.

$$\Pr(\mathcal{N}(r) = k) \leq 2 \exp\left(-\frac{\pi r^2 n_s p}{L^2}\right) \frac{\left(\frac{\pi r^2 n_s p}{L^2}\right)^k}{k!}.$$

Proof: Let $\mathcal{A}(d, r_1, r_2)$ denote the overlapping area of two circles of radius r_1, r_2 with centers of distance d away, and $\mathcal{C}(r)$ denotes number of nodes within radius r . Note that the distribution of cluster clients is HPP within a circle of radius R , thus we can obtain

$$\begin{aligned} \Pr(\mathcal{N}(r) = k) & = \sum_{m=0}^{n_s} \Pr(\mathcal{N}(r) = k | \mathcal{C}(r+R) = m) \Pr(\mathcal{C}(r+R) = m) \\ & = \sum_{m=0}^{n_s} \left(\int_0^{R+r} \sum_S \left(\prod_{i=1}^m (e^{-\mu_1} \frac{\mu_1^{v_i}}{v_i!}) \right) \frac{2x}{(R+r)^2} dx \right) e^{-\mu_2} \frac{\mu_2^m}{m!} \\ & = \int_0^{R+r} \left(\sum_{m=0}^{n_s} e^{-m\mu_1} \frac{(m\mu_1)^k}{k!} e^{-\mu_2} \frac{(\mu_2)^m}{m!} \right) \frac{2x}{(R+r)^2} dx \\ & \leq \int_0^{R+r} \left(e^{-\mu_1 \mu_2} \frac{(\mu_1 \mu_2)^k}{k!} \right) \frac{2x}{(R+r)^2} dx \\ & = \int_0^{R+r} e^{-\frac{\mathcal{A}(x, R, r) n_s p}{L^2}} \frac{\left(\frac{\mathcal{A}(x, R, r) n_s p}{L^2}\right)^k}{k!} \frac{2x}{(R+r)^2} dx. \end{aligned}$$

During the above derivation, we utilize the following notations to simplify our calculations:

$$\begin{aligned} \mu_1 & = \frac{\mathcal{A}(x, R, r) p}{\pi(R+r)^2}, \mu_2 = \frac{\pi(R+r)^2 n_s}{L^2} \text{ and } \sum_{i=1}^m v_i = k \\ \mathcal{S} & = \{v_0, v_1 \dots v_m | \sum_{i=1}^m v_i = k\} \end{aligned}$$

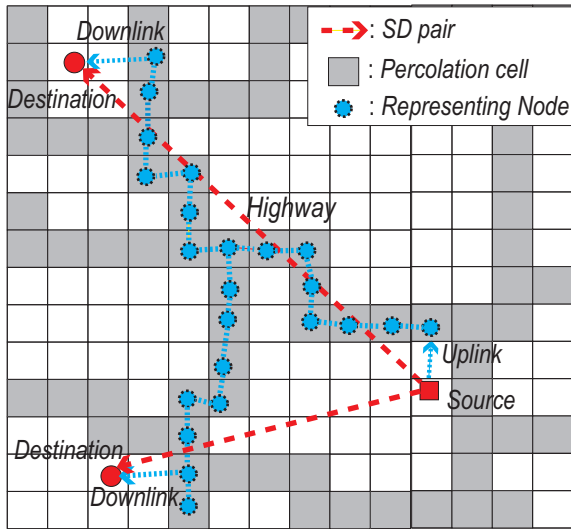


Fig. 4: Demonstration of routing protocol.

The overlapping area of two circles is smaller than the area of the small circle therefore

$$A(x, R, r) \leq \begin{cases} \pi r^2 & x \in [0, R + r) \\ 0 & x \in [R + r, \infty) \end{cases}$$

Then we substitute it into $\Pr(\mathcal{N}(r) = k)$ and obtain

$$\Pr(\mathcal{N}(r) = k) \leq 2e^{-\frac{\pi r^2 n_s p}{L^2}} \frac{(\frac{\pi r^2 n_s p}{L^2})^k}{k!}.$$

Lemma 5.2: There exists a constant τ such that if we equally partition the square \mathcal{O} into cells with side length $\tau L / \sqrt{n_s p}$, the probability that at least one node resides in a cell is larger than $1 - 2 \exp(-\frac{\pi \tau^2}{4})$.

Proof: Each cell of edge length $\tau L / \sqrt{n_s p}$ contains a disk of radius $\frac{\tau L}{2\sqrt{n_s p}}$ such that the probability that at least one node resides in a cell can be lower bounded as follows:

$$\Pr(\mathcal{N}(r) \geq 1) \geq 1 - 2 \exp(-\frac{\pi r^2 n_s p}{L^2}) = 1 - 2 \exp(-\frac{\pi \tau^2}{4}).$$

Now we outline the definition of an information highway: Choose τ large enough so that $1 - 2 \exp(-\frac{\pi \tau^2}{4}) > 5/6$ and we equally partition the square \mathcal{O} into cells of edge $\tau L / \sqrt{n_s p}$. Thus there are $\lfloor \frac{\sqrt{n_s p}}{\tau} \rfloor \times \lfloor \frac{\sqrt{n_s p}}{\tau} \rfloor$ cells and each cell in i th row and j th column is denoted by $s_{i,j}$. $s_{i,j}$ is open if it contains at least one node. A horizontal (vertical) cross path is defined as a set of open cells that cross \mathcal{O} from left to right (top to bottom). The gray cell in Figure 4 is example of percolation path. According to Theorem 5 in [3], if the probability that a cell contains at least 1 node is larger than $\frac{5}{6}$, there are w.h.p. $\Theta(n_s p)$ disjoint paths crossing from left to right (top to bottom). A set containing these $\Theta(n_s p)$ horizontal and vertical paths are called *information highways*. Note that the cells composed of the highway are called percolated cell, and the nodes are called representing nodes. The following properties

of information highways are proved in [3] via percolation theory.

- In each horizontal (vertical) rectangle of size $L \times (\kappa \log(n_s p) \frac{L}{\sqrt{n_s p}} - \epsilon_L)$, there are at least $\delta \log(n_s p)$ horizontal (vertical) highway paths w.h.p. It indicates there are $\Theta(\sqrt{n_s p})$ disjoint crossing paths from left to right and top to bottom, respectively. (Theorem 5)
- The length of each crossing path is bounded by $\Theta(L)$.
- The distance between two adjacent horizontal (vertical) paths is at most $O(L \log(n_s p) / \sqrt{n_s p})$.
- There exists a spatial and temporal scheme that can achieve $\Theta(W)$ throughput on the highway. It means that each cell composing the highway can be considered linked by optical wires to its neighbors with bandwidth $\Theta(W)$. (Theorem 3)

The information highway is an infinitely large component such that each node in the deployed region can connect to it within a hop of length $O(\frac{L \log(n_s p)}{\sqrt{n_s p}})$. Now we need to construct a multicast spanning tree \mathcal{T}_j for \mathcal{C}_j to route its packets.

Lemma 5.3: The Prim's algorithm is utilized to construct an \mathcal{T}_j for cluster \mathcal{C}_j and we prove that $|\mathcal{T}_j| \leq 6\sqrt{2}|\mathcal{C}_j|R$.

Proof: Prim's algorithm: Initially, each node is a separate part, then we iteratively find the shortest edge to compose a larger part until one part is left. Each member in \mathcal{C}_j is confined in a disk of radius R . We utilize a square of edge $2R$ to cover the whole circle. At each i th ($1 \leq j \leq |\mathcal{C}_j|$) step, there are $|\mathcal{C}_j| + 1 - i$ parts remaining. We equally partition the square into $\lfloor \sqrt{|\mathcal{C}_j| + 1 - i} \rfloor^2$ cells with edge length $\frac{2R}{\lfloor \sqrt{|\mathcal{C}_j| + 1 - i} \rfloor}$, and there exists at least one cell which contains more than 2 parts, which means the shortest edge connecting two parts in i th step is at most $\frac{2\sqrt{2}R}{\lfloor \sqrt{|\mathcal{C}_j| + 1 - i} \rfloor}$. Therefore, the upper bound of $|\mathcal{T}_j|$ is:

$$\begin{aligned} |\mathcal{T}_j| &\leq \sum_{i=1}^{|\mathcal{C}_j|} \frac{2\sqrt{2}R}{\lfloor \sqrt{|\mathcal{C}_j| + 1 - i} \rfloor} \\ &= 2\sqrt{2}R \left(\lfloor \sqrt{|\mathcal{C}_j|} \rfloor + \sum_{i=1}^{\lfloor \sqrt{|\mathcal{C}_j|} \rfloor} \frac{1}{i} + \frac{|\mathcal{C}_j|}{\lfloor \sqrt{|\mathcal{C}_j|} \rfloor} - 2 \right) \\ &\leq 6\sqrt{2}|\mathcal{C}_j|R. \end{aligned} \quad (16)$$

Based on the above analysis, a capacity routing scheme is provided in Algorithm 2.

According to Algorithm 2, the average number of nodes a percolated cell has to serve is $\frac{\kappa}{\delta}$. The next lemma illustrates the minimum achievable data rate in the uplink and downlink phases.

Lemma 5.4: In the uplink and downlink phase, a data rate of $\Omega(\log^{-2}(n_s p))$ can be sustained for each transmission.

³ κ and δ is some constant, $\epsilon_L = o(\log(n_s p) \frac{L}{\sqrt{n_s p}})$ and is to make $\kappa \log(n_s p) \frac{L}{\sqrt{n_s p}} - \epsilon_L$ an integer.

Proof: According to Algorithm 2, for each node X , both $|\mathcal{F}_h(X) - X|$ and $|\mathcal{F}_v(X) - X|$ is upper bounded by $\kappa \log(n_{sp}) \frac{L}{\sqrt{n_{sp}}}$. It means that if we equally divide the region \mathcal{O} into **sub-squares** of size $2\kappa \log(n_{sp}) \frac{L}{\sqrt{n_{sp}}} \times 2\kappa \log(n_{sp}) \frac{L}{\sqrt{n_{sp}}}$, X then $\mathcal{F}_h(X)$ must reside in the same cell. The same is true for X and $\mathcal{F}_v(X)$. Then according to Lemma 6 in [11], each cell can be allocated a constant time to be active within the mini-slot. According to Lemma 3.2, there are at most $8\kappa^2 \log^2(n_{sp})$ nodes in each cell, **which means** there are at most $8\kappa^2 \log^2(n_{sp})$ **SD pair**. Thus, each **SD** pairs can be allocated $\frac{1}{O(\log^2(n_{sp}))} = \Omega(\log^{-2}(n_{sp}))$ fraction of a time slot for transmission. ■

Algorithm 2 Capacity Achieving Scheme for UCRM

- 1: **Access Point Mapping:** Establish mappings $\mathcal{F}_h(X)$ and $\mathcal{F}_v(X)$ for each node X . Horizontally divide the $L \times L$ square into slices of size $L \times (\kappa \log(n_{sp}) \frac{L}{\sqrt{n_{sp}}} - \epsilon_L)$. Then there are at least $\delta \log(n_{sp})$ paths in each slice. Denote each path in the slice as $path_i (1 \leq i \leq \delta \log(n_{sp}))$. We further divide each slice into $\delta \log(n_{sp})$ **sub-slices** of size $L \times (\frac{\kappa L}{\delta \sqrt{n_{sp}}})$ each. If node X is in the i_{th} sub-slice, $\mathcal{F}_h(X)$ denotes the percolated cell on $path_i$ with the same ordinate. **Mapping of $\mathcal{F}_v(X)$ is the dual of $\mathcal{F}_h(X)$, and is mapped by applying the above algorithm to vertical paths.**
 - 2: **Medium Access in Highway:** Each representing node can be active for a constant portion of time in a cell partitioned network based on Lemma 6 in [11]. Therefore there exists a spatial and temporal accessing policy such that each representing node can deliver $O(W)$ bits to its adjacent representing node as [3].
 - 3: **Routing Protocol:** A multicast spanning tree \mathcal{T}_j is constructed in Lemma 5.3. Each time slot is divided into 3 mini-slot and corresponds to 3 phases as in Figure 4. For each branch in \mathcal{T}_j linking nodes X_1 and X_2 , the 3 phases are as follows:
 - **Uplink:** X_1 drains its data to $\mathcal{F}_h(X_1)$.
 - **Highway:** This phase corresponds to step 2 and **utilizes** multihop transmission along the horizontal path from $\mathcal{F}_h(X_1)$ to the intersection with the vertical path in which $\mathcal{F}_v(X_2)$ resides then **forwards** to $\mathcal{F}_v(X_2)$.
 - **Downlink:** X_2 downloads the data from $\mathcal{F}_v(X_2)$.
-

Then we begin to analyze the second phase. **Although** the highway can be virtually considered **as** a wired network with bandwidth $\Theta(W)$, each path must help relay data for **many** clusters. Therefore the allocated radio resources for a cluster is limited. In the following part, we will study the maximum number of clusters a percolated cell **can serve**.

Lemma 5.5: Each percolated cell cannot relay data for cluster with head at a distance of $\sqrt{2}(R + \frac{2\kappa \log(n_{sp})\tau L}{\sqrt{n_{sp}}})$ away from the cell, therefore $\mathcal{R} \leq \sqrt{2}(R + \frac{2\kappa \log(n_{sp})\tau L}{\sqrt{n_{sp}}})$.

Proof: We refer to Figure 5 for the proof. Each path is constrained within a strip of width $\kappa \log(n_{sp}) \frac{L}{\sqrt{n_{sp}}}$. Thus D is upper bounded by $(\kappa \log(n_{sp}) \frac{L}{\sqrt{n_{sp}}})$. **Recalling** that the radius

of the disk is R , we know the farthest cell that could be used is the black cell, a distance $\mathcal{R} \leq \sqrt{2}(R + \frac{(1+\kappa \log(n_{sp}))\tau L}{\sqrt{n_{sp}}}) \leq \sqrt{2}(R + \frac{2\kappa \log(n_{sp})\tau L}{\sqrt{n_{sp}}})$ away from the kernel. ■

Each branch in the spanning tree \mathcal{T}_j is regarded as a SD pair linking two nodes. The length of it determines how many hops should be used on the highway.

Lemma 5.6: Assume that the length of a SD pair is ℓ , then the number of hops required is $\frac{c_3}{\tau} \left(\frac{\sqrt{2n_{sp}\ell}}{L} + 4\kappa \log(n_{sp}) \right)$, where c_3 is a constant.

Proof: Denote the Abscissa and Ordinate of a point $\xi \in \mathcal{O}$ as ξ_a and ξ_o . Assume ξ^1 and ξ^2 are two representing points in two percolated cells on the same horizontal path. Let $h(\xi^1, \xi^2)$ denote the number of hops required for transmission from ξ^1 to ξ^2 . We can prove that there exists a constant c_3 , such that

$$\Pr \left(h(\xi^1, \xi^2) \leq \frac{c_3 \sqrt{n_{sp}} |\xi_a^1 - \xi_a^2|}{\tau L} \right) \geq 1 - \delta(n_{sp}), \quad (17)$$

where $\lim_{n_{sp} \rightarrow \infty} \delta(n_{sp}) = 0$. Now we divide the path into $K = \frac{L}{|\xi_a^1 - \xi_a^2|}$ sub-paths as in the second figure in Figure. 5 and $\ell_i = \frac{L \varphi(n_{sp}, i)}{K} (1 \leq i \leq K)$ denotes the length of the i_{th} sub-path. Then the proof of Eqn. (17) is **as follows**.

$$\lim_{n_{sp} \rightarrow \infty} \Pr(\ell_i = \omega(\frac{L}{K})) = 0,$$

which is also identical to prove

$$\lim_{n_{sp} \rightarrow \infty} \Pr(\varphi(n_{sp}, i) = \infty) = 0.$$

For each path on the highway system, the length of the **path** is on the order of $\Theta(L)$ according to its property. Thus there exists a constant c' , such that the length of every path is upper bounded by $c'L$, which means $\sum_{i=1}^K \ell_i \leq c'L$. Taking the **expected value** on both sides, we can obtain

$$\begin{aligned} & \sum_{i=1}^K \inf_{\varphi(n_{sp}, i) = \infty} \{ \varphi(n_{sp}, i) \} \Pr(\ell = \omega(L/K)) \frac{L}{K} \\ & \leq \sum_{i=1}^K \mathbf{E}[\ell_i] = \mathbf{E}[\sum_{i=1}^K \ell_i] \leq c'L. \end{aligned}$$

Then we can obtain

$$\Pr(\ell = \omega(L/K)) \leq \frac{c'}{\inf_{\varphi(n_{sp}, i) = \infty} \{ \varphi(n_{sp}, i) \}}$$

Taking $\delta(n_{sp}) = \frac{c'}{\inf_{\varphi(n_{sp}, i) = \infty} \{ \varphi(n_{sp}, i) \}}$ and note that each hop can forward the packet of distance $\Theta(L/\sqrt{n_{sp}})$ on the highway, we complete the proof of Eqn. (17). The vertical path is the dual of the horizontal case.

Now assume that a sender at ξ^s generates traffic to a receiver ξ^r at a distance ℓ away. Let ξ^i be the intersection of the horizontal and vertical path in which $\mathcal{F}_h(\xi^s)$ and $\mathcal{F}_v(\xi^r)$ reside. According to the properties of percolation **paths**, every percolation path is constrained within a strip of

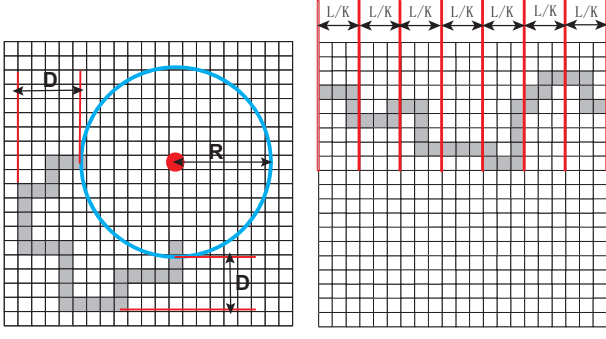


Fig. 5: Demonstration of influential range and a division of a percolation path

width $\kappa \log(n_s p) \frac{L}{\sqrt{n_s p}}$ such that

$$\begin{cases} \max\{|\mathcal{F}_h(\xi^s)_a - \xi_a^i|\} \leq |\xi_a^s - \xi_a^i| + \frac{\kappa L \log(n_s p)}{\sqrt{n_s p}} \\ \max\{|\mathcal{F}_v(\xi^r)_o - \xi_o^i|\} \leq |\xi_o^r - \xi_o^i| + \frac{\kappa L \log(n_s p)}{\sqrt{n_s p}} \\ \max\{\sqrt{(\xi_a^s - \xi_a^i)^2 + (\xi_o^r - \xi_o^i)^2}\} \leq \ell + \frac{2\kappa L \log(n_s p)}{\sqrt{n_s p}} \end{cases}$$

Then the number of hops required for transmission with nodes ℓ away is upper bounded by

$$\begin{aligned} & h(\mathcal{F}_h(\xi^s), \xi^i) + h(\xi^i, \mathcal{F}_v(\xi^r)) \\ & \leq \frac{c_3 \sqrt{n_s p}}{\tau L} (|\mathcal{F}_h(\xi^s)_a - \xi_a^i| + |\mathcal{F}_v(\xi^r)_o - \xi_o^i|) \\ & \leq \frac{c_3}{\tau} \left(\frac{\sqrt{n_s p}}{L} (|\xi_a^s - \xi_a^i| + |\xi_o^r - \xi_o^i|) + 2\kappa \log(n_s p) \right) \\ & \leq \frac{c_3}{\tau} \left(\frac{\sqrt{n_s p}}{L} \sqrt{2((\xi_a^s - \xi_a^i)^2 + (\xi_o^r - \xi_o^i)^2)} + 2\kappa \log(n_s p) \right) \\ & \leq \frac{c_3}{\tau} \left(\frac{\sqrt{2n_s p} \ell}{L} + 4\kappa \log(n_s p) \right). \end{aligned}$$

Lemma 5.7: Assume that cell s is within a distance \mathcal{R} from \mathcal{C}_j 's head. **Then** if $R \geq \frac{\kappa \log(n_s p) L}{3\sqrt{n_s p}}$, the probability \mathcal{P} that s is utilized to transmit for cluster \mathcal{C}_j is upper bounded as:

$$\mathcal{P} \leq \frac{12c_3 \delta \tau L \sqrt{|\mathcal{C}_j|}}{\kappa \sqrt{n_s p} R}. \quad (18)$$

Proof: There are at least $\frac{\kappa}{\delta} \lfloor \frac{\sqrt{2n_s p} \mathcal{R}}{\tau L} \rfloor^2 \geq \frac{\kappa}{\delta} (\frac{\sqrt{2n_s p} \mathcal{R}}{\tau L} - 1)^2$ percolated cells within the disk of radius \mathcal{R} . Follow the prim algorithm used in Lemma 5.3 to construct a spanning tree. Let $\mathcal{I}(s, i)$ be the indicator whether cell s is used in the i th step and \mathcal{P} be the probability that s is used in the whole process.

Then based on Lemma 5.5, 5.6:

$$\begin{aligned} \mathcal{P} & \leq \sum_{j=1}^{|\mathcal{C}_j|} \Pr(\mathcal{I}(s, i) = 1) \\ & \leq \sum_{j=1}^{|\mathcal{C}_j|} \frac{c_3}{\tau \frac{\kappa}{\delta} (\frac{\sqrt{2n_s p} \mathcal{R}}{\tau L} - 1)^2} \left(\frac{4\sqrt{n_s p} R}{\lfloor \sqrt{|\mathcal{C}_j|} + 1 - i \rfloor L} + 4\kappa \log(n_s p) \right) \\ & \leq \frac{c_3 \delta \tau L^2}{2\kappa n_s p R^2} \sum_{j=1}^{|\mathcal{C}_j|} \left(\frac{4\sqrt{n_s p} R}{\lfloor \sqrt{|\mathcal{C}_j|} + 1 - i \rfloor L} + 4\kappa \log(n_s p) \right) \\ & \leq \frac{c_3 \delta \tau L^2}{2\kappa n_s p R^2} \left(\frac{12\sqrt{n_s p} |\mathcal{C}_j| R}{L} + 4\kappa |\mathcal{C}_j| \log(n_s p) \right) \\ & \leq \frac{12c_3 \delta \tau L \sqrt{|\mathcal{C}_j|}}{\kappa \sqrt{n_s p} R}. \end{aligned}$$

The above lemma provides an upper bound on the probability that a multicast flow from a cluster \mathcal{C}_j is routed via the cell when it is within the influential range of that cluster. Let us now study the total number of times that the cell is used by our capacity achieving scheme. We know that $|\mathcal{C}_j| = n^{1-\alpha_i}$ if $\mathcal{C}_j \in \mathcal{SC}_i$. Let $\mathcal{P}_i \leq \frac{12c_3 \delta \tau L \sqrt{n^{1-\alpha_i}}}{\kappa \sqrt{n_s p} R} \leq \frac{12c_3 \delta \tau L}{\kappa \sqrt{n_s p} R}$ denote the probability that a cell is used for \mathcal{C}_j if $\mathcal{C}_j \in \mathcal{SC}_i$ and s is within the influential range \mathcal{R} of \mathcal{C}_j 's head. For each cluster \mathcal{C}_j , $|\mathcal{C}_j|$ clients are randomly and independently distributed in a disk of radius R . And each SD pair generated in the Prim's Algorithm is also a random process. Then we will apply Vapnik-Chervonenkis theorem to prove our results and the VC-dimension of a multicast spanning tree is $O(\log p)$ according to [4].

Theorem 5.1: (VC-Theorem): If \mathcal{S} is a set of finite VC-dimension $\text{VC-d}(\mathcal{S})$, and $\{X_i | i = 1, 2, \dots, N\}$ is a sequence of i.i.d. random variables with common probability distribution P , then for every $\epsilon, \delta > 0$,

$$\Pr \left(\sup_{A \in \mathcal{S}} \left| \frac{\sum_{i=1}^N I(X_i \in A)}{N} - P(A) \right| \leq \epsilon \right) > 1 - \delta$$

when $N \geq \max \left\{ \frac{8\text{VC-d}(\mathcal{S})}{\epsilon} \log \frac{13}{\epsilon}, \frac{4}{\epsilon} \log \frac{2}{\delta} \right\}$.

Theorem 5.2: For each \mathcal{SC}_i ($1 \leq i \leq k$), let $k(1) = 1$ and $k(i) = 1/k$ for $2 \leq i \leq k$, then the following inequality should be satisfied based on VC-Theorem.

$$\Pr \left(\sup_{s \in \mathcal{O}} \left(\mathcal{FL}_i(s) \leq \frac{24\pi c_3 \delta \tau}{\kappa} \frac{R n^\alpha}{k(i) L \sqrt{n^{\alpha_i}}} \right) \right) \geq 1 - \delta(n),$$

where $\mathcal{FL}_i(s)$ is the number of flows using s by clusters belonging to \mathcal{SC}_i and $\delta(n)$ is a set of sequences approaching 0 when n goes to infinity. Eqn. (21) should be satisfied when $2 \leq i \leq k$ and Eqn. (22) should be satisfied when $i = 1$.

Proof: The method to construct multicast trees are i.i.d. variables within a disk of radius R . Then the multicast trees can be viewed as i.i.d variables. Therefore, we can use the VC-Theorem. Recalling that, given a cell s , the probability that a multicast tree will cross it is at most $\frac{12c_3 \delta \tau L \sqrt{|\mathcal{C}_j|}}{\kappa \sqrt{n_s p} R}$ when the

cell is within its influential range \mathcal{R} . Hence, the whole set of the cell \mathcal{O} :

$$\Pr \left(\sup_{s \in \mathcal{O}} \left| \frac{\mathcal{F}\mathcal{L}_i(s)}{N} - \mathcal{P}_i \right| \leq \epsilon(n) \right) > 1 - \delta(n)$$

$$\text{when } N \geq \max \left\{ \frac{8d}{\epsilon(n)} \log \frac{13}{\epsilon(n)}, \frac{4}{\epsilon(n)} \log \frac{2}{\delta(n)} \right\},$$

where $d = O(\log p)$ is the VC-dimension. Then substitute Eqn. (18) into it and note that $|\mathcal{C}_j| = \Theta(n^{1-\alpha_i})$ and we obtain

$$\Pr \left(\sup_{s \in \mathcal{O}} \frac{\mathcal{F}\mathcal{L}_i(s)}{N} \leq \frac{12c_3\delta\tau L}{\kappa\sqrt{n^{\alpha_i}R}} + \epsilon(n) \right) > 1 - \delta(n). \quad (19)$$

Now let $\epsilon(n) = \frac{12c_3\delta\tau L}{\kappa\sqrt{n^{\alpha_i}R}}$ and $\delta(n) = \frac{2}{n}$ and for inequality (19) to be true,

$$\begin{aligned} N &\geq \max \left\{ \frac{8d}{\epsilon(n)} \log \frac{13}{\epsilon(n)}, \frac{4}{\epsilon(n)} \log \frac{2}{\delta(n)} \right\} \\ &= \frac{\kappa R n^{\alpha_i/2} \log n}{3c_3\delta\tau L}, \end{aligned} \quad (20)$$

is required. In this case, N denotes the number of clusters **belonging** to $\mathcal{S}\mathcal{C}_i$ ($2 \leq i \leq k$) within the disk of radius R . Applying the same technique as Lemma 3.2, we can obtain that $N \in (\frac{\pi n_s R^2}{2kL^2}, \frac{2\pi n_s R^2}{kL^2})$ if $R = \omega(\sqrt{\frac{k}{n_s}}L)$. Therefore when $N \geq \frac{\pi n_s R^2}{2kL^2} \geq \frac{\kappa R n^{\alpha_i/2} \log n}{3c_3\delta\tau L}$, if

$$R \geq \frac{2\kappa k L n^{\alpha_i/2} \log n}{3\pi c_3 \delta \tau n_s} = \frac{2\kappa}{3\pi c_3 \delta \tau} k L n^{\alpha_i/2-\alpha} \log n, \quad (21)$$

Eqn. (20) is satisfied. Substitute $N \leq \frac{2\pi n_s R^2}{kL^2}$ into Eqn. (19) and we complete the proof when $2 \leq i \leq k$. A similar technique is employed for $\mathcal{S}\mathcal{C}_1$ if the following condition is satisfied.

$$R \geq \frac{2\kappa L n^{\alpha_i/2} \log n}{3\pi c_0 c_3 \delta \tau n_s} = \frac{2\kappa}{3\pi c_0 c_3 \delta \tau} \frac{L \log n}{\sqrt{n^\alpha}}. \quad (22)$$

The above constraint of R means that only when R is **sufficiently** large, the number of flows over a certain cell s can be upper bounded. In addition, we find that if the number of flows over s in $\mathcal{S}\mathcal{C}_{i'}$ is upper bounded by Δ , Δ is also an upper bound of the number of flows in $\mathcal{S}\mathcal{C}_i$ ($i' \leq i \leq k$) **because of** α_i is an increasing sequence over i .

Lemma 5.8: Given a cluster radius $R = \omega(L/\sqrt{n_s})$, the data rate that the highway can sustain for \mathcal{C}_j is denoted by λ_j^h . When $R = O(\sqrt{\frac{k}{n_s}}L)$,

$$\lambda_j^h = \Omega \left(\frac{L}{R\sqrt{n^\alpha}} \right) \quad 1 \leq j \leq k,$$

when $R = \omega(\sqrt{\frac{k}{n_s}}L)$,

$$\lambda_j^h = \begin{cases} \Omega \left(\frac{L n^{\alpha_i/2}}{R n^\alpha} \right) & \mathcal{C}_j \in \mathcal{S}\mathcal{C}_i, \quad 1 < i \leq k', \\ \Omega \left(\frac{L n^{\alpha_{k'}/2}}{R n^\alpha} \right) & \mathcal{C}_j \in \mathcal{S}\mathcal{C}_i, \quad k' < i < k. \end{cases} \quad (23)$$

Proof: In case of $R = O(\sqrt{\frac{k}{n_s}}L)$, there are at most $O(\frac{n_s R^2}{L^2}) \leq O(k)$ clusters within a disk of radius R therefore only a small portion of super clusters may have members inside the disk. The number of flows through each **cell** is upper bounded by $\Theta(\frac{R\sqrt{n_s}}{L})$ w.h.p according to Theorem 5.2. Therefore $\hbar = O(\frac{R\sqrt{n_s}}{L})$ in our FDM network and the highway can sustain a data rate of $W/\hbar = \Omega(\frac{L}{R\sqrt{n_s}})$.

When $R = \omega(\sqrt{\frac{k}{n_s}}L)$, the number of clusters **belonging** to $\mathcal{S}\mathcal{C}_i$ ($2 \leq i \leq k$) is $\Theta(\frac{n_s R^2}{kL^2})$ and $\hbar = k$ which means each $\mathcal{S}\mathcal{C}_i$ ($2 \leq i \leq k$) is allocated $(1-c_0)W/k$ bandwidth and $\mathcal{S}\mathcal{C}_1$ is allocated c_0W bandwidth for transmission in the highway. We specify a k' as follows:

$$k' = \arg \sup_{1 \leq i \leq k} \left\{ \frac{2\kappa}{3\pi c_0 c_3 \delta \tau} k L n^{\alpha_i/2-\alpha} \log n \leq R \right\}. \quad (24)$$

Then we know the number of flows across each percolation cell s can be upper bounded by $\frac{24\pi c_3 \delta \tau}{\kappa} \frac{R n^\alpha}{k L \sqrt{n^{\alpha_i}}}$ for clusters **belonging** to an arbitrary super cluster $\mathcal{S}\mathcal{C}_i$ ($1 \leq i \leq k'$). Therefore

$$\lambda_j^h \geq \frac{W/k}{\frac{24\pi c_3 \delta \tau}{\kappa} \frac{R n^\alpha}{k L \sqrt{n^{\alpha_i}}}} = \frac{\kappa W L \sqrt{n^{\alpha_i}}}{24\pi c_3 \delta \tau R n^\alpha} = \Omega \left(\frac{L \sqrt{n^{\alpha_i}}}{R n^\alpha} \right).$$

For $\mathcal{C}_j \in \mathcal{S}\mathcal{C}_i$ ($k' + 1 \leq i \leq k$), the number of flows **crossing** a cell s is also upper bounded by $\frac{24\pi c_3 \delta \tau}{\kappa} \frac{R n^\alpha}{k L \sqrt{n^{\alpha_{k'}}}}$, **which is the same as that of clusters belonging to** $\mathcal{S}\mathcal{C}_{k'}$ and in this case $\lambda_j^h = \Omega \left(\frac{L n^{\alpha_{k'}/2}}{R n^\alpha} \right)$. ■

Recalling Lemma 5.4, we know the bottleneck is due to data delivery on the highway when $R \geq \Omega \left(\frac{L \log n}{\sqrt{n^\alpha}} \right)$. Let λ_j^u, λ_j^d denote the data rate of the uplink and downlink, respectively, and $\lambda_j = \min\{\lambda_j^u, \lambda_j^d, \lambda_j^h\} = \lambda_j^h$. The achievable capacity λ for the whole network is therefore

$$\lambda = \min\{\lambda_j | 1 \leq j \leq n_s\} = \Omega \left(\frac{L}{R\sqrt{n^\alpha}} \right) = \Omega \left(\frac{L}{R\sqrt{n_s}} \right).$$

Therefore our scheme approaches the **same maximum** capacity as Eqn. (15) in order sense if we substitute $R = \frac{L}{\sqrt{\pi(1+L^2\sigma\sigma)}}$ into λ and $\lambda = \Omega(\max\{\frac{1}{\sqrt{n_s}}, \frac{L\sigma\sigma}{\sqrt{n_s}}\})$.

VI. DISCUSSION

Our results can generalize various results obtained in non-heterogeneous network like [3], [4], [6]–[8]. Here we highlight two features in our clustered network, *heterogeneous cluster traffic* and *heterogeneous cluster size*, both of which are characterized in **real networks** with multicast traffic **patterns**.

A. The Impact of Heterogeneous Cluster Traffic

HCT characterizes a network property that clients of each multicast session are not uniformly distributed across the network. In many cases, the packets are more likely to be delivered to adjacent nodes. **Previous** works are insufficient to estimate the achievable capacity. Indeed, Eqn. (5) is a precise formula for the achievable capacity. However, we cannot decide whether HCT increases the network capacity because there is no criteria to judge the level of heterogeneity

for a dispersion density function. By introducing the variable *distribution variance* $\sigma_{\mathcal{O}}$, we offer a quantitative description of the extent of heterogeneous traffic. We find that if the traffic of a cluster is not uniformly disseminated across the region as is assumed in prior works, the network capacity λ increases.

Figure 6 illustrates the relationship between λ and $\sigma_{\mathcal{O}}$ derived from Eqn. (15). In region 1, HCT is **relatively** slight and each cluster is **full overlap** with other clusters. Fully overlapping indicates that in each small region $\mathcal{O}' \subseteq \mathcal{O}$ with area $\Theta(L^2/n)$, there are $\Theta(n_s) = \Theta(n^\alpha)$ clusters whose members have approximately equal probability to reside in \mathcal{O}' . Therefore it is much **like** the uniform distributed cases and the network capacity is not improved. In region 2, HCT begins to influence the network performance by increasing the maximized capacity. Here clusters are partially overlapped, it indicates that in each small region $\mathcal{O}' \subseteq \mathcal{O}$ with area $\Theta(L^2/n)$, there are nearly $\Theta(n^\theta)$ clusters whose members have approximately equal probability to reside in \mathcal{O}' . In this case, $\theta < \alpha$ and θ is **closely** related to σ . And the implicit reason for an increased capacity is that each relay only needs to deliver packets for a smaller portion of the clusters compared to **the** previous one. In region 3, trivial overlapping means in each small region $\mathcal{O}' \subseteq \mathcal{O}$ with area $\Theta(L^2/n)$, there are at most $\Theta(1)$ clusters whose members have approximately equal probability to reside in \mathcal{O}' . Therefore each relay only needs to deliver packets for a constant number of clusters and the achievable capacity tends **towards** $O(W)$.

B. The Impact of Heterogeneous Cluster Size

HCS characterizes a network property that each multicast session **is not comprised of the same number of clients**. We show that HCS is not a deterministic factor for the achievable network capacity λ , **although individual capacities among clusters may differ**. Based on Eqn. (23), we provide a **rough** estimation of λ_j : $\lambda_j = \Theta(\sqrt{\frac{n^{\alpha_i}}{n^\alpha}} \lambda)$ if $\mathcal{C}_j \in \mathcal{SC}_i$. The relationship between capacity and cluster size is illustrated in Figure 7. During our derivation, we propose a new concept of super clusters, which **contain** clusters of the same size in order sense. The reason why we propose such a concept is that the scaling property of the achievable capacity is sustained by a certain group. It is more convenient to handle a discrete variable than a continuous one and we can gain a general insight **into** the impact of the cluster size. **Note** that we have only offered a point in the whole capacity region to show that smaller clusters can have a much larger capacity than the network capacity. **Measuring the entire capacity region is beyond the scope of this paper.**

The capacity achieving scheme is only for UCRM in this work. Such a property can be generalized to other network **topologies** as well. In addition, a strong assumption is made in our paper that the number of **clusters** with size $\Theta(n^{1-\alpha}) = \Theta(p)$ must be larger than $c_0 n_s$. Such an assumption is crucial for highway construction. **We** will discuss some more general HCS in our future work.

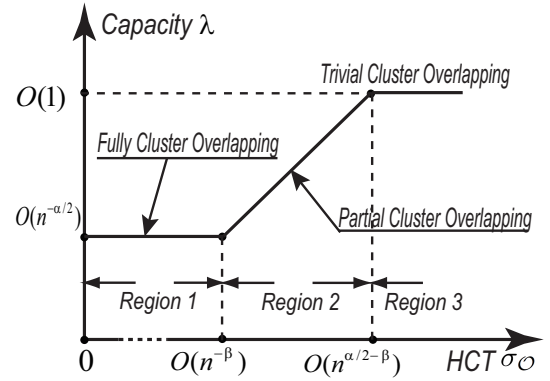


Fig. 6: Relationship between capacity λ and $\sigma_{\mathcal{O}}$ (log scale). In region 2, the network capacity increases with $\sigma_{\mathcal{O}}$. $\sigma_{\mathcal{O}} = 0$ means a non-heterogeneous network with the same achievable capacity as prior works.

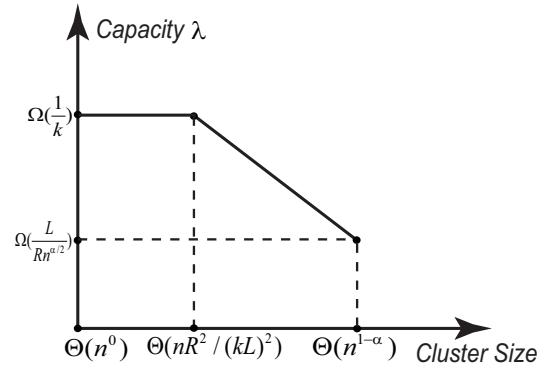


Fig. 7: Relationship between capacity λ and Cluster Size (log scale).

VII. CONCLUSION AND FUTURE WORK

In this paper, we study the effect of heterogeneity on the asymptotic multicast capacity in clustered **networks**. Our contributions are mainly divided into two parts. First, we find that heterogeneous cluster traffic increases the achievable capacity for all the clusters. Through analyzing the fundamental constraints of wireless **networks** from global and local aspects, a quantitative **relationship** is provided between network capacity λ and *distribution variance* $\sigma_{\mathcal{O}}$. **This is the first time this relationship has been used to describe heterogeneity in literature. The effect of heterogeneous cluster size is analyzed in uniform cluster random model, which is the optimal network layout given a fixed $\sigma_{\mathcal{O}}$.**

We find it cannot **increase** λ but increase the achievable capacity for small clusters under our framework.

There are several **tasks for us to complete** in the future. First, **the cluster dense model should be studied. The cluster dense model indicates that the distance between cluster heads approaches 0 when $n \rightarrow \infty$.** Second, the dispersion density function $\phi(\cdot)$ for each cluster is assumed identical. **It would be interesting to study networks with clusters of different**

$\phi(\cdot)$. Finally, the assumption of heterogeneous cluster size is a little restrictive here so a more general analysis is crucial for obtaining more insight into heterogeneous cluster size.

REFERENCES

- [1] P. Gupta, P. R. Kumar. "The capacity of wireless networks," IEEE Trans. on Info. Theory, Vol. 46, No. 2, pp. 388-404, Mar.2000.
- [2] L. L. Xie and P. R. Kumar, "A network information theory for wireless communication: Scaling laws and optimal operation " IEEE Trans. Info. Theory, vol. 50, pp. 748-767, 2004.
- [3] M. Franceschetti, O. Dousse, D.N.C. Tse, P. Thiran. "Closing the gap in the capacity of random wireless networks via percolation theory," IEEE Trans. on Info. Theory, Vol. 53, No. 3, pp. 1009-1018, 2007.
- [4] X.-Y. Li, S.-J. Tang, and O. Frieder. "Multicast capacity for large scale wireless ad hoc networks," in Proc. ACM Mobicom'08, San Francisco, CA, USA, Sept. 2008.
- [5] A. Keshavarz-Haddad, V. Ribeiro, R. Riedi, "Broadcast capacity in multihop wireless networks," in Proc. ACM Mobicom'06, Los Angeles, CA, USA, 2006.
- [6] S. Shakkottai, X. Liu, and R. Srikant. "The multicast capacity of large multihop ad hoc networks," in Proc. ACM Mobihoc'07, Montreal, Quebec, Canada, 2007.
- [7] S. Li, Y. Liu, X.-Y. Li, "Capacity of Large Scale Wireless Networks Under Gaussian Channel Model," in Proc. ACM MobiCom'08, San Francisco, CA, USA, Sept. 2008.
- [8] C. Wang, X.-Y. Li, C. Jiang, S. Tang, Y. Liu, J. Zhao. "Scaling Laws on Multicast Capacity of Large Scale Wireless Networks," in Proc. IEEE INFOCOM'09, Rio de Janeiro, Brazil, April 2009.
- [9] C. Hu, X. Wang, F. Wu, "MotionCast: On the Capacity and Delay Tradeoffs," in ACM MobiHoc'09, New Orleans, USA, May 2009.
- [10] S. Zhou and L. Ying, "On Delay Constrained Multicast Capacity of Large-Scale Mobile Ad-Hoc Networks," in Proc. INFOCOM'10 mini-conference, San Diego, CA, USA, March 2010.
- [11] X. Lin, Ness B. Shroff, "The Fundamental Capacity-Delay Tradeoff in Large Mobile Wireless Networks," in Proc. IEEE MedHoc'04, Bodrum, Turkey, June 27-30, 2004.
- [12] L. Ying, S. Yang R. Srikant. "Optimal Delay-Throughput Trade-Offs in Mobile Ad Hoc Networks," IEEE Tran. on Info. Theory, Vol. 54, No. 9, September, 2008.
- [13] U. Niesen, P. Gupta, D. Shah, "The Multicast Capacity Region of Large Wireless Networks," in Proc. of IEEE INFOCOM'09, Rio de Janeiro, Brazil, April 2009.
- [14] C. Hu, X. Wang, D. Nie, J. Zhao, "Multicast Scaling Laws with Hierarchical Cooperation," in Proc. of IEEE INFOCOM'10, San Deigo, CA, USA, 2010.
- [15] G. Alfano, M. Garetto, E. Leonardi, "Capacity Scaling of Wireless Networks with Inhomogeneous Node Density: Upper Bounds," IEEE Journal on Selected Areas of Communications, Vol. 27, No. 7, September 2009.
- [16] G. Alfano, M. Garetto, E. Leonardi, "Capacity Scaling of Wireless Networks with Inhomogeneous Node Density: Lower Bounds," in Proc. IEEE INFOCOM'09, Rio de Janeiro, Brazil, April 2009.
- [17] Z. Wang , H. R. Sadjadpour and J. J. Garcia-Luna-Aceves, "A Unifying Perspective on the Capacity of Wireless Ad Hoc Networks," in Proc. IEEE INFOCOM'08, Phoenix, AZ, USA, April 2008.
- [18] M. Ji, Z. Wang, H. Sadjadpour, J. J. Garcia-Luna-Aceves, "The Capacity of Ad Hoc Networks with Heterogeneous Traffic Using Cooperation," in Proc. IEEE INFOCOM'10, San Diego, CA, USA, March, 2010.
- [19] P. Li and Y. Fang, "The Capacity of Heterogeneous Wireless Networks," in Proc. IEEE INFOCOM'10, San Diego, CA, USA, March, 2010.
- [20] D.-Z. Du and F.-K. Hwang. "A proof of the gilbert-pollak conjecture on the steiner ratio," in Algorithmica 7((2,3)):121-135, 1992.
- [21] J. M. Steele. "Growth rates of euclidean minimal spanning trees with power weighted edges." in The Annals of Probability, 16(4):1767-1787, Oct 1988.
- [22] Q. Peng, X. Wang, H. Tang, "Heterogeneity Increases Multicast Capacity in Clustered Network," in Proc. of IEEE INFOCOM 2011, Shanghai, China.
- [23] M. D. Penrose, "The Longest Edge of the Random Minimal Spanning Tree," The Annals of Applied Probability, vol. 7, no. 2, pp. 340-361, 1997.