Scheduling Algorithms and Bounds for Rateless Data Dissemination in Dense Wireless Networks

Kan Lin, David Starobinski, Ari Trachtenberg
Department of Electrical and Computer Engineering
Boston University
Boston, Massachusetts, USA
Email: {linkan,staro,trachten}@bu.edu

Sachin Agarwal
NEC Europe
Heidelberg, Germany
Email: ska@alum.bu.edu

***Abstract*—Many applications in wireless cellular networks rely on the ability of the network to reliably and efficiently disseminate data to a large client audience. The stochastic nature of packet loss across receivers and channel interference constraints between cells complicate this task, however. In this paper, we analyze the problem of minimizing the delay of data dissemination in dense multi-channel wireless cellular networks, using rateless coding transmission. We begin with an extreme value analysis of the delay in a single cell setting, and show that the growth rate of this random variable becomes deterministic as the client audience scales up. Next, we extend the analysis to multi-cell, multi-channel settings and derive tight performance bounds on the delay. Our analysis reveals that the availability of more channels does not always reduce delay proportionally. This sub-linear gain effect is guaranteed to occur if the difference between the chromatic number and the fractional chromatic number of the graph is greater than one.**

I. Introduction

The proliferation of affordable mobile devices has significantly boosted deployments of wireless cellular networks (e.g., Wi-Fi, 3G, LTE, etc.). Such wireless cellular networks typically consist of two components: base stations, which provide backbone communication, and wireless clients; these are illustred in Fig. 1 with a five-cell wireless network, where each base station (henceforth *clusterhead*) is surrounded by clients within its radio range (collectively called a *cluster*).

The client population within one cluster can be quite high for modern wireless networks; for example, the statistics bureau of Japan reports a population density of 5751 persons per sq km in Tokyo [1], many of whom carry smartphones. Indeed, growing density in wireless networks is an inevitable trend, which in turn, raises challenges for some data transmission applications.

This paper focuses on the fundamental problem of designing efficient data dissemination strategies in such dense wireless networks, where communication occurs through a broadcast channel, as may be witnessed, for example, when smartphones receive updates of events happening around their location or when wireless sensors schedule periodic firmware updates. The aim of data dissemination is to broadcast a set of identical data packets from a media source, to all targeted wireless clients in the network through the clusterheads in an efficient manner (i.e., short time duration). Several intrinsic factors of wireless networks make this objective particularly challenging.

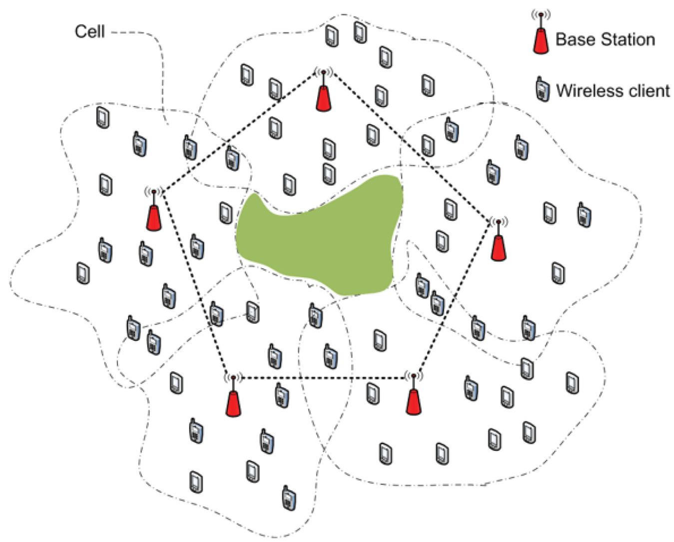

Fig. 1. Illustration of a wireless cellular network. Wireless clients are connected to base stations (clusterheads), which are interconnected via reliable wireline/wireless links themselves. The dotted irregular shapes contour radio signal range of each clusterhead. It is assumed that one or more of the clusterhead is connected to greater networks (not shown in the figure).

The first difficulty rises from high rates of *packet losses* typically experienced by wireless clients. Given a big client audience, base stations can be overwhelmed by the number of retransmission requests, a phenomenon called "broadcast storm" [2], making traditional (N)ACK-based retransmission mechanism infeasible. Instead, a *rateless code*, which is a form of packet-wise forward error correction (FEC) scheme, allow each wireless client to decode packets from *any* sufficiently large subset of encoded packets. Yet, there still remains the question of how many broadcasting packets are needed for the data dissemination to complete, and especially how this number is affected by the stochastic environment of lossy wireless channels.

The second difficulty results from the diversity of wireless network topologies. This difficulty emerges mainly because of the radio channel conflicts among clusters: simultaneous operation on the same frequency by neighboring clusterheads will cause interference and can affect data packet reception. As such, broadcasting must be carefully scheduled to avoid this type of interference problem, potentially reducing dissemination efficiency.

The third difficulty arises when one wishes to exploit channel diversity; for example, in the U.S., IEEE 802.11 wireless nodes can transmit over 11 different channels in the 2.4 GHz band, three of which are non-overlapping. In an effort to maximize dissemination efficiency while avoiding interference, one needs to solve the allocation problem of these channel resources over both space and time.

Our goal is to determine optimal strategies for minimizing the time required for all the wireless clients to receive a disseminated file (i.e. maximizing dissemination efficiency) for general multi-cell topologies and multiple channels, given all the factors described above, using rateless coding transmission.

Our contributions are as follows. First, we conduct a study of the fluctuation of the completion time in a single cluster. Specifically we perform an extreme value analysis of the completion time, asymptotic in increasing client audience N, and find that the completion time scales proportionally to $\log(N) + o(\log(N))$ in probability. This result implies that, despite the stochastic nature of packet loss, the problem lends itself to a *deterministic* optimization problem. We further argue that this deterministic phenomenon also spreads out to multiple clusters, which substantiates the feasibility of employing a deterministic scheduling policy. Thus, we propose asymptotically optimal transmission policies for general multi-cluster, multi-channel topologies and provide tight lower and upper bounds on their completion time. As part of our analysis, we find that dissemination efficiency is subject to a *sub-linear gain* with respect to growing channel availability, unlike the case for cluster chains [3] where linear gain can be achieved. This sublinear curve stands as the absolute bound that any asymptotical scheduling policy can achieve, a fact that is valuable for those who design and operate cellular networks.

The rest of the paper is organized as follows: Section II discusses related work and introduces graph coloring background that is necessary to our later analysis. In Section III we describe our models of the problem and its formulation. Section IV deals with single cluster rateless coding transmission, where the fluctuation of dissemination delay is studied and its deterministic trend is revealed. We expand our analysis to general multiple cluster topologies in Section V with details of our optimal policy and performance bounds. Finally, we conclude this paper in Section VI.

II. Preliminaries

A. Related Work

Considerable effort has been invested in the literature in search of solutions to some of the issues described above.

In [4] the authors conduct a quantitative analysis on the performance of FEC codes, such as rateless codes, in the presence of a large client population. An asymptotic cumulative distribution function (CDF) of the completion time is derived. However, this work only considers the analysis of a single cluster, whereas our work consider the fluctuation of the completion time itself for its asymptotic behavior, and take multiple clusters organized along an arbitrary topology into consideration as well.

The work in [3] analyzes plaintext data dissemination over simple cluster topologies, namely, linear chains of clusters. The limitation of this work is that it does not consider the scenario of rateless coding, nor general multi-cluster topologies. Moreover, unlike our work, the results of [3] apply to the expected completion time, while our work consider the behavior of the completion time random variable itself.

The work in [5] analyzes the gain of network coding (another FEC coding scheme) in a layered, multi-transmitter multi-receiver systems, where packets are passed over layer by layer using a random transmission scheme. However, this model is inconsistent with ours, where packets are considered to be ready in all base stations at first and spread out simultaneously instead of being passed down layer by layer.

There has been substantial work on resource allocation for maximizing the throughput of multi-hop wireless networks (see [6] for a survey). In [13] the authors show that maximizing throughput for a single channel over one-hop is equivalent to the classical Maximum Weighted Matching problem, solvable in polynomial time. The problem of obtaining an efficient interference-aware link scheduling scheme for a wireless network is also considered in [12]. The authors in [14] present low-complexity (although not optimal) algorithms for distributed scheduling of wireless nodes in a network. In [7] the authors study the multi-flow problem and showed the NP-hardness of finding Maximum Independent Set (MIS).

Our paper is distinct from this previous body work in several aspects. First, we consider delay as the primary optimization metric rather than throughput. Since network traffic is typically bursty, the delay of a burst of data (generically referred to as a *file* in this paper) is often the primary metric of interest. Second, we explicitly model packet loss rather than assuming that communication is loss free even if certain interference constraints are satisfied. This leads to a more realistic model since wireless links are notoriously prone to packet losses due to fading, etc. Finally, our analysis focuses on the case of densely populated networks, the key fact that significantly simplifies the design of scheduling algorithms for lossy wireless networks, as detailed in the sequel.

B. Graph Coloring Background

Our analysis in this paper is related to some fundamental graph coloring problems, the most significant of which we list herein for sake of completeness.

Definition 1: An *independent set* of a graph G is a set of vertices, such that no two of them are adjacent. A *maximal independent set (MIS)* of G is an independent set that is not a subset of any other independent set of G.

We use a matrix $\mathbf{I}$ to represent a collection of independent sets (or maximal independent sets) of a graph $G(V, E)$. Each row index corresponds to one vertex and each column represents a distinct independent set. Every element in this matrix takes on a value of either 0 or 1, with 1 indicating the existence of the corresponding vertex in the associated independent set.

Definition 2: A *vertex coloring* of a graph G is a labeling of the graph's vertices with colors such that no adjacent vertices have the same color. The *chromatic number* $\chi(G)$ of G is the smallest number of colors among all vertex coloring scheme of the graph.

Definition 3: A *b-fold coloring* of a graph G is an assignment of color sets of size b to vertices, such that each vertex is assigned one set and adjacent vertices receive disjoint sets. An $a : b$-coloring is a b-fold coloring, whose sets are subsets of a universe of a available colors. The *b-fold chromatic number* χ_b(G) of G is the smallest a for which an $a : b$-coloring exists.

Definition 4: The *fractional chromatic number* $\chi_f(G)$ is defined to be

$$\chi_f(G) = \lim_{b\to\infty} \frac{\chi_b(G)}{b} = \inf_b \frac{\chi_b(G)}{b}.$$

The difference between $\chi_f(G)$ and $\chi(G)$ is unbounded in general. More details about b-fold and fractional coloring problems can be found in [9].

III. Models and Problem Formulation

A. Network Topology

We assign the scattered wireless clients (also simply called *nodes*) to a set of *clusters* based on physical proximity to base stations (also called *clusterheads*), as shown in Fig. 1. Specifically, all the nodes associated with a given base station are considered as belonging to the this cluster. A node is associated with only one base station, as is typically the case in practice, even if it is under the range of multiple base stations. Base stations transmit data packets via broadcast, that is, each packet is simultaneously transmitted to all the nodes within radio range.

Physical interference constraints induce a logical graph $G(V, E)$, where the vertex set V represent base stations, and edge set E represent interference constraints i.e., an edge between vertex i and j means that the the coverage areas (clusters) of base stations i and j overlap. For example, the cluster graph corresponding to Fig. 1 is a pentagon, one cluster per node. We assume that the cluster topology is arbitrary but fixed, i.e., it does not change during data dissemination.

We also assume that cluster i contains $N_i = \alpha_i N$ nodes, where α_i is a constant coefficient, with $\sum_{i=1}^{|V|} \alpha_i = 1$, and N refers to the scaling variable. In this paper we focus on the asymptotic regime of a large client population. Therefore, our analysis applies to the case where $N \to \infty$.

B. Transmission Model

We consider the problem of disseminating a file consisting of M packets to all the wireless clients via broadcasts from base stations. We assume that links between base stations are fast and reliable, meaning that the entire file is available at every base station before data dissemination proceeds. The time axis is slotted and each broadcast packet transmission takes one time slot. Base stations are equipped with a single radio, as is typically the case, meaning that a node can either transmit or receive (but not both) on one channel at a point of time.

Furthermore, we consider the use of *rateless coding* for transmitting the file. Instead of transmitting the original M packets, the base station transmits a long sequence of encoded packets (e.g., using random linear codes) of these M packets. In each slot, the base station generates and transmits a new packet. The generating algorithm is devised in such a way that as soon as a node receives M different correct packets[1], it can restore the original file.

C. Channel Interference Model

We assume that there are C non-overlapping channels available in total that can be operated on by base stations, all of them having identical, but independent statistical characteristics. The packet loss probability on each channel is denoted p. Thus, the probability that a node within a cluster does not correctly receive a packet broadcast by the base station is p, independently of all other events (e.g., correct reception or not of that packet by other nodes in the cluster).

In general, the presence of interference or crosstalk between neighboring clusters results in jamming. We assume that a packet transmission in a cluster on a specific channel will bring interference to its neighboring clusters if they communicate on the same channel, i.e., all cluster nodes will be unable to correctly receive packets. As a result, transmission on a channel in a cluster bars all its neighboring clusters from operating on the same channel during the same time slot.

D. Problem Formulation

Our objective henceforth is to determine a *control policy* that minimizes the completion time (i.e., the total number of transmission time slots needed) until all the N nodes receive the entire file (M distinct decoded packets) in a network with C channels and interference constraints modeled by a graph G. We denote $T_N^C(G)$ the minimum completion time obtained using the optimal control policy.

IV. Single Cluster Topology

In this section, we investigate the performance of data dissemination in single cluster topology as the first step to develop analysis for general multi-cluster topologies. Specifically, we consider a single cluster containing N nodes. Since there is no interference constraints from neighboring clustering in this case, we do not expect benefit of using multiple channels. Hence, we simplify the notation of completion time in this section to T_N.

Because of the fact that rateless coding is designed to handle multiple clients, the optimal control policy of this case is nothing but letting the clusterhead keep transmitting new rateless encoded packets until completion. In the rest of this section, we analyze the completion time of data dissemination T_N under this policy and show that its growth becomes

[1]This is true for an *ideal* rateless code. In practical scenarios, a receiver may need slightly more packets for decoding.

deterministic for large N. Our analysis use the standard asymptotic $O(\cdot)$ and $o(\cdot)$ notations [11].

Consider first the completion time for a specific node n ($1 \leq n \leq N$), which we denote $T(n)$. The probability that the completion time for this node takes exactly $M + k$ slots is given by a negative binomial distribution.

The CDF of this probability distribution can be expressed as $\Pr\{T(n) \leq M + k\} = 1 - I_p(k+1, M)$, where $I_p(\cdot)$ corresponds to the so-called *regularized incomplete beta* function [10]:

$$I_p(k+1, M) = \sum_{j=k+1}^{k+M} \frac{(k+M)!}{j!(k+M-j)!} p^j (1-p)^{k+M-j}. \quad (1)$$

Given N nodes, the completion time of the entire cluster is the *maximum* value of the completion time of all its nodes, namely, $T_N = \max\{T(1), T(2), \cdots, T(N)\}$.

Next, we examine how T_N scales as N grows. The relationship between them is described with the help of a function $k(N)$. When considering a specific function $k(N)$, the following lemma, whose proof is omitted, provides us with a sufficient condition on the convergence of probabilities.

Lemma 1: If $\lim_{N\to\infty} N \cdot I_p(k(N)+1, M) = 0$, then $\lim_{N\to\infty} \Pr\{T_N \leq M + k(N)\} = 1$. On the other hand, if $\lim_{N\to\infty} N \cdot I_p(k(N)+1, M) = \infty$, then $\lim_{N\to\infty} \Pr\{T_N \leq M + k(N)\} = 0$.

As a result, the following theorem predicts the completion time of data dissemination.

Theorem 1: For fixed M and p, the completion time of data dissemination using rateless coding in a single cluster is, with probability one,

$$T_N = \log_\lambda N + o(\log_\lambda N),$$

where $\lambda = 1/p$. In other words, $\frac{T_N}{\log_\lambda N} \to 1$ in probability.

Proof: The proof follows from bounding the expression $N \cdot I_p(k(N)+1, M)$ followed by showing that the choices of $k(N) = \log_\lambda N + M \log_\lambda \log_\lambda N$ and $k(N) = \log_\lambda N - \log_\lambda \log_\lambda N$ respectively make $N \cdot I_p(k+1, M)$ go to 0 (in the plus sign case) and 1 (in the minus sign case). The proof is fulfilled by applying this result to Lemma 1. ■

The key insight offered by this theorem is that, despite the stochastic nature of the environment, the order of growth of the completion time becomes deterministic, when N gets large. We conducted simulation of the delay for parameters $M = 10, p = 0.2$ over 4500 iterations and found that 35% of instances occurs at the mode of 21 packets when $N = 10^3$, while 43% of instances occurs at the mode of 29 packets when $N = 10^7$. Further, in both cases, over 90% of the samples have a completion time that differ at most by two packets from the mode. This property greatly simplifies the design of efficient scheduling policies for multiple clusters, as explained in the next section.

V. Multiple Cluster Topologies

In this section, we extend our analysis to the general case of multiple clusters with multiple channels. As described in Section III, we use the notation of $N_i = \alpha_i N$ where $\sum_i^{|V|} \alpha_i = 1$ to represent the number of nodes in the i-th cluster. As a result, when considering the transmission demand for an individual cluster, the results of Section IV remains applicable by replacing N with $\alpha_i N$.

We first show that asymptotically the same time is needed for each cluster to complete dissemination to all nodes. Our problem is then reduced to a scheduling problem among the clusters. Our primary goal is to determine the optimal control policy and derive its performance, and secondary goal is to assess the impact of increasing the number of channels available for data dissemination on the completion time performance. All the asymptotic results of this section hold with probability one as $N \to \infty$.

A. Structure of the Optimal Transmission Policy

Denote T_{N_i} to be the total number of transmissions required for cluster i. Though these numbers differ amongst different clusters for finite N, they scale identically as N gets large, as the following lemma (whose proof is also omitted) demonstrates, based on Theorem 1.

Lemma 2: For a fixed set of coefficients α_i, $i = 1, 2, \cdots, |V|$, the total number of transmission slots required by different clusters scale identically as $N \to \infty$, namely

$$T_{N_i} = \log_\lambda N + o(\log N), \qquad \forall \quad 1 \leq i \leq |V|.$$

This lemma shows that as $N \to \infty$ the *transmission demand* of different clusters becomes identical to T_N, the number given by Theorem 1.

According to our channel interference model, we further have the following lemma (proof is omitted).

Lemma 3: For any transmission policy that induces interference, there exists another policy that does not induce interference and causes no greater transmission delay.

In summary, to solve the problem of finding an asymptotically optimal transmission policy, it is sufficient to find for each time slot an appropriate maximal independent set (MIS) for each channel. Note that the selected MIS may change from one time slot to the next.

B. Analysis

In this section, we derive the optimal policy for general multi-cluster topologies with multiple channels. Then, we provide bounds on the optimal completion time that reveal an interesting relationship with both $\chi(G)$ and $\chi_f(G)$.

1) Optimal Transmission Policy: The availability of multiple channel resource allows us to select C MISs for each time slot (a different channel is assigned to each MIS). As a result, some clusters may belong to multiple MISs and have multiple channels available for transmission, whereas only one of them can be used.

We recall that $\mathbf{I}$ corresponds to be the matrix of all r maximal independent sets of G. We also introduce a matrix $\mathbf{P}$ in which each row index i represents a different MIS (corresponding to a different column of $\mathbf{I}$) and each column j represents a different *transmission schedule* corresponding to a different combination of C MISs chosen out of r available. Therefore the dimensions of this matrix are $r \times \binom{r}{C}$. Note that the sum of the elements in each column of $\mathbf{P}$ is equal to C. An entry $\mathbf{P}_{ij}$ is set to 1 if MIS i belongs to transmission schedule j.

Note that we assume $r > C$, since the case of $C \leq r$ is trivial: each MIS can be assigned a channel so the completion time is the same as for a single cluster.

The product $\mathbf{I} \cdot \mathbf{P}$ yields a $|V| \times \binom{r}{C}$ matrix. The value of the (i, j) entry of this matrix corresponds to the number of channels allocated to cluster i using transmission schedule j. Since each cluster can only use one channel in a given time slot, we introduce a cutoff function $f(\cdot)$, which takes an integer matrix as its argument and returns a binary matrix of the same dimension.

$$f(\mathbf{A})_{ij} = \begin{cases} 1 & \text{if } \mathbf{A}_{ij} \geq 1 \\ 0 & \text{if } \mathbf{A}_{ij} = 0. \end{cases}$$

In this way, $f(\mathbf{I} \cdot \mathbf{P})$ gives the number of effective channels usable in each cluster (i.e., 0 or 1).

Our optimization problem is thus:

$$\begin{aligned} &\text{Minimize} && \sum_{i=1}^{\binom{r}{C}} x_i \\ &\text{Subject to} && f\left(\begin{bmatrix} & \mathbf{I} & \end{bmatrix} \cdot \begin{bmatrix} & \mathbf{P} & \end{bmatrix}\right) \begin{bmatrix} x_1 \\ \vdots \\ x_{\binom{r}{C}} \end{bmatrix} \geq \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \\ &&& x_1, \cdots, x_{\binom{r}{C}} \geq 0. \end{aligned} \tag{2}$$

where x_i ($i = 1, \cdots, x_{\binom{r}{C}}$) stands for the portion of time that the network should transmit using a specific transmission schedule, i.e., a specific combination of C MISs out of r available as indicated by the i-th column of matrix $\mathbf{P}$. For convenience, we denote $\mathbf{x}$ and $\mathbf{1}$ the column vector $[x_1, \cdots, x_{\binom{r}{C}}]^T$ and $[1, \cdots, 1]^T$ respectively.

We denote $\theta(G, C)$ the optimal objective value returned by (2), i.e., $\theta(G, C) = \min(\sum_{i=1}^{\binom{r}{C}} x_i)$. This number refers to the amount of time normalized by T_N. Clearly, $\theta(G, C)$ is no less than 1. The minimal dissemination delay achievable is $T_N^C(G) = \theta(G, C) T_N$ by applying the following policy:

- Solve the programming problem (2).
- Transmit in MISs indicated by the j-th column $\mathbf{P}$ for $x_j / \sum_{i=1}^{\binom{r}{C}} x_i$ portion of time, until every cluster finishes. The order does not matter.

2) Performance bound: We now bound $\theta(G, C)$. The purpose of these bounds is to assess the gain achieved by the availability of additional channels and establish a relationship between $\theta(G, C)$ and the chromatic and fractional chromatic numbers of the graph.

Computing the optimal policy involves solving programming problem (2). To compute a lower bound, we first replace $f(\mathbf{I} \cdot \mathbf{P})$ by $\mathbf{I} \cdot \mathbf{P}$ in the constraint, namely,

$$\begin{aligned} &\text{Minimize} && \sum_{i=1}^{\binom{r}{C}} x_i \\ &\text{Subject to} && \mathbf{I} \cdot \mathbf{P} \cdot \mathbf{x} \geq \mathbf{1} \\ &&& \mathbf{x} \geq 0. \end{aligned} \tag{3}$$

Because $f(\mathbf{I} \cdot \mathbf{P}) \leq \mathbf{I} \cdot \mathbf{P}$ by definition of $f(\cdot)$, the solution of (2) also satisfies all constraints of (3) but does not necessarily achieve the minimum of (3). Denoting $\theta'(G, C)$ the solution of (3), we have $\theta'(G, C) \leq \theta(G, C)$.

Using this relaxation, the following lemma shows the completion time for C channels is at most C times smaller than that for a single channel if $C \leq \chi_f(G)$. The proof of this lemma is omitted for space consideration.

Lemma 4: For $C \leq \chi_f(G)$

$$\theta(G, C) \geq \frac{\chi_f(G)}{C}.$$

Note that this bound is tight when $C = 1$.

The following lemma shows that the (normalized) completion time for C channels is smaller or equal to $\chi(G)/C$ if $C \leq \chi(G)$. Moreover, by the definition of the chromatic number, this bound is tight when $C = \chi(G)$. The proof of this lemma is omitted for space consideration.

Lemma 5: For $C \leq \chi(G)$,

$$\theta(G, C) \leq \frac{\chi(G)}{C}.$$

Combining Lemmas 4 and 5, we obtain the following bounds on the performance of the optimal policy in general topologies with multiple channels:

Theorem 2:

$$\max\left(\frac{\chi_f(G)}{C}, 1\right) \leq \theta(G, C) \leq \max\left(\frac{\chi(G)}{C}, 1\right)$$

or equivalently,

$$\max\left(\frac{\chi_f(G)}{C} T_N, T_N\right) \leq T_N^C(G) \leq \max\left(\frac{\chi(G)}{C} T_N, T_N\right).$$

3) Sub-linear performance gain: Theorem 2 shows that the completion time decreases with the number of channels. With one channel the lower bound in Theorem 2 can always be achieved. However, with increasing channels this lower bound is no longer guaranteed. Generally, for topologies whose fractional coloring number $\chi_f(G)$ differs from the vertex coloring number $\chi(G)$, performance will diverge from the lower bound $\chi_f(G)/C$ as C increases. We refer to this phenomenon as to the *sub-linear gain effect*. Theorem 2 implies a sufficient condition on C for this type of sub-linear gain effect to occur, namely $\chi_f(G) < C < \chi(G)$. This result holds because we know that if $\chi(G)/C > 1$ then $\theta(G, C) \geq 1$ (since $\theta(G, C) = 1$ only when $C \geq \chi(G)$). Note that the condition $\chi_f(G) < C < \chi(G)$ is sufficient but not necessary for the sub-linear effect to occur as shown by the next example.

To illustrate, we consider a specific random graph G of 12 clusters (not shown for space consideration), with fractional

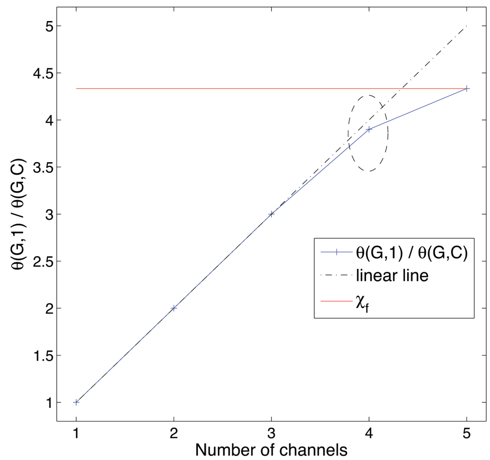

Fig. 2. The performance gain ratio $\theta(G, 1)/\theta(G, C)$ of a random graph of 12 clusters. The horizontal line indicates the highest (linear) gain achievable. The case $C = 4$ (see circled region) is where the sub-linear gain effect occurs.

chromatic number $\chi_f(G) = 4.333$ and chromatic number $\chi(G) = 5$. By solving (2), we obtain $\theta(G, C)$ for different values of C. As it turns out, the optimal performance for $C = 1$ is $\theta(G, 1) = \chi_f(G) = 4.333$. For $C = 2$ and $C = 3$, we respectively have $\theta(G, 2) = \chi_f(G)/2 = 2.167$ and $\theta(G, 3) = \chi_f(G)/3 = 1.444$. However, for $C = 4$, we have $\theta(G, 4) = 1.111 > \chi_f(G)/4 = 1.083$. Ultimately, for $C = 5$, $\theta(G, 5) = \chi(G)/5 = 1$. Fig. 2 depicts the ratio of $\theta(G, 1)$ to $\theta(G, C)$ and illustrates the divergence of the performance gain from the linear line when $C = 4$.

The sub-linear effect results from the nature of rateless coding, i.e., having multiple channels available in a given cluster does not help in reducing the completion time performance of its member nodes. Thus, the cases of linear performance gain for small C are achieved by going to great lengths to search for alternative MISs, in an effort to balance channel coverage. The sub-linear effect occurs when such other MIS options fail.

The occurrence rate of the sublinear gain effect greatly depends on the network size (in terms of clusters) and interference constraints. In fact, we conducted a simulation using 1000 Erdős-Rényi random graphs (probability of connection between any two vertices is set 0.3). When the network size is $|V| = 10$ clusters, less than 2% of all graphs showed sub-linear gain effect, but when $|V|$ increases to 16 clusters, this percentage increased to close to 15%.

VI. Conclusion

We consider the problem of achieving minimum delay rateless data dissemination in lossy, multi-channel wireless cellular networks with arbitrary topologies. We show that the transmission demand (i.e., the number of transmissions required in each cluster) grows almost deterministically when the number of clients grows to infinity. This enables the use of linear programming to provide asymptotically optimal transmission policies.

Our analysis shows that the completion time in a single channel multi-cluster topology using the optimal policy is about χ_f larger than the completion time in a single cluster, where χ_f is the fractional chromatic number of the graph representing the inter-cluster topology. As the number of channels made available for data dissemination is increased, the completion time is reduced until the number of channels is equal to χ, the chromatic number of the graph. Since the fractional chromatic number is always smaller or equal to the chromatic number, the gain in completion time grows less than linearly in the number of channels. More precisely, we show that if there exists a value C such that $\chi_f < C < \chi$, then the sub-linear effect is guaranteed to occur. Thus, this condition always holds if $\chi - \chi_f > 1$. These results could be useful for network providers who are gauging the usefulness of acquiring additional spectrum for their network.

Acknowledgment

This work was supported in part by the National Science Foundation under grant CCF-0916892 and a grant from Deutsche Telekom Laboratories. This work was completed when Sachin Agarwal was employed by Deutsche Telekom Laboratories.

References

[1] "Statistical Handbook of Japan 2010", The Statistics Bureau and the Director-General for Policy Planning of Japan, 2010.

[2] Y-C Tseng et al., "The broadcast storm problem in a mobile ad hoc network", Wireless Networks, vol. 8, no. 2/3, 2002

[3] D.Starobinski and W.Xiao, "Asymptotically Optimal Data Dissemination in Multichannel Wireless Sensor Networks: Single Radios Suffice" IEEE/ACM Transactions on Networking, vol. 18, no.3, pp. 695-707, Jun. 2010.

[4] W. Xiao and D. Starobinski, "Extreme Value FEC for Reliable Broadcasting in Wireless Networks," IEEE JSAC (Special Issue on Simple Wireless Sensor Networking Solutions), Vol. 28, No. 7, pp. 1180-1189, Sept. 2010.

[5] A. Eryilmaz, A. Ozdaglar, M. Medard, E. Ahmed , "On the delay and throughput gains of coding in unreliable networks", IEEE Transactions on Information Theory, vol.54, no.12, pp. 5511–5524, Dec. 2008.

[6] X. Lin, N. B. Shroff, and R. Srikant, "A tutorial on cross-layer optimization in wireless networks", IEEE Journal on Selected Areas in Communications, vol. 24, no. 8, pp. 1452–1463, Aug. 2006.

[7] P. Wan, "Multiflows in multihop wireless networks", Proceedings of the tenth ACM international symposium on Mobile ad hoc networking and computing, New Orleans, Louisiana, May. 2009.

[8] David J. C. MacKay, "Information Theory, Inference, and Learning Algorithms", Cambridge University Press, ISBN 0-521-64298-1, 2003.

[9] E. R. Scheinerman, D. H. Ullman, "Fractional Graph Theory", Wiley, 1997

[10] Frank W. Olver, Daniel M. Lozier, Ronald F. Boisvert, et.al., "NIST Handbook of Mathematical Functions", Cambridge University Press, New York, 2010.

[11] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, "Introduction to Algorithms", Second Edition, The MIT Press, Sep. 2001.

[12] W. Wang, et.al., "Efficient Interference-Aware TDMA Link Scheduling for Static Wireless Networks", ACM MOBICOM 2006, Los Angeles, California, USA, Sep. 2006.

[13] G. Sharma, R. Mazumdar, and N. B. Shroff, "On the complexity of Scheduling in Wireless Networks", ACM MOBICOM 2006, Los Angeles, California, USA, Sep. 2006.

[14] A. Gupta, X. Lin, and R. Srikant, "Low-Complexity Distributed Scheduling Algorithms for Wireless Networks", IEEE INFOCOM 2007, Anchorage, Alaska, USA, 2007.