Wireless Broadcast with Network Coding: A Connected Dominating Sets Approach

Cédric Adjih, Song Yean Cho

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Wireless Broadcast with Network Coding: A Connected Dominating Sets Approach

Cédric Adjih, Song Yean Cho

N° 6547

Juin 2008

Thème COM

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--6547--FR+ENG

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

INRIA

centre de recherche
PARIS - ROCQUENCOURT

Wireless Broadcast with Network Coding: A Connected Dominating Sets Approach

Cédric Adjih, Song Yean Cho

Thème COM — Systèmes communicants
Équipe-Projet Hipercom

Rapport de recherche n° 6547 — Juin 2008 — 30 pages

Abstract: We study network coding for multi-hop wireless networks. We focus the case of broadcasting, where one source transmits information to all the nodes in the network. Our goal is energy-efficient broadcasting, in other words, to minimize the number of transmissions for broadcasting to the entire network. To achieve this goal, we propose a family of methods that combine the use of network coding and connected dominating sets. They consists in rate selections using connected dominated sets (RAUDS: Rate Adjustment Using Dominating Sets, and an generalized version, MARAUDS). The main insight behind these methods is that their use of connected dominating sets, allows near-optimality in the core of the network, while they efficiently handle borders and non-uniformity. The main contribution is a formal proof of the performance of these families of algorithms. One main result is the comparison of performance between routing and these methods (and in general, network coding).

Key-words: wireless networks, network coding, broadcasting, multi-hop, min-cut, hypergraph, connected dominating set

Centre de recherche INRIA Paris – Rocquencourt
Domaine de Voluceau, Rocquencourt, BP 105, 78153 Le Chesnay Cedex
Téléphone : +33 1 39 63 55 11 — Télécopie : +33 1 39 63 53 30

Diffusion dans les réseaux sans fil avec le codage réseau: une approache utilisant les ensembles dominants connectés

Résumé : Nous étudions le codage réseau pour les réseaux sans fil multi-sauts. Nous nous intéressons au cas de la diffusion, où une source transmet des informations à tous les noeuds du réseau. Notre objectif est une diffusion efficace en énergie, c'est-à-dire, qui vise à minimiser le nombre de transmissions faites pour diffuser à tout le réseau. Pour parvenir à cet objectif, nous proposons une famille de méthodes qui combinent l'utilisation du codage réseau et les ensembles dominants connectés. Elles consistent en une selection de débit, utilisant des ensembles connectés dominants (RAUDS: "Rate Adjustment Using Dominating Sets", et une version généralisée, MARAUDS). L'idée principale sous-jacente à ces méthodes, est que leur utilisation des ensembles dominants, permet une d'opérer de manière quasi-optimale dans le coeur du réseau, et en même temps d'être capable de traiter les problèmes aux frontières du réseaux, et la non-uniformité. Un résultat principal est aussi une comparaison de la performance entre le routage et ces méthodes (et, plus généralement, le codage réseau).

Mots-clés : réseaux sans fil, codage de réseau, diffusion, multi-sauts, coupe minimale, hypergraphe, ensemble dominant connecté

Contents

1 Introduction

Seminal work from Ahlswede, Cai, Li and Yeung in [1] has introduced the idea of *network coding*, whereby intermediate nodes are mixing information from different flows (different bits or different packets).

In multi-hop wireless networks, one of natural application of network coding is to reduce the number of transmissions required to transmit some amount of information to the same destinations. This achieves energy-efficiency for networks where the cost of wireless communication is a critical design factor. We focus on one specific form of communication, broadcasting information from one source to all the nodes in a wireless multi-hop network. Then energy efficiency may be formulated as:

- Given one source, minimize the total number of (re)transmissions used to allow all nodes in the network to get the information.

This issue of efficient broadcast in multi-hop wireless networks has been traditionally addressed by a rich literature including methods such as connected dominating sets (CDS) for instance. Examples of CDS algorithms include the algorithms from Guha and Khuller [2], the protocol from Das et Bharghavan [3] or the localized method of Dai and Wu [4]. In these approaches, every node is a neighbor of at least one node in the CDS: broadcast may be performed by having each node in the CDS retransmitting one source packet exactly once. The efficiency of one algorithm is directly given by the number of nodes inside the CDS.

The issue of efficient multicast (hence broadcast) has also been studied with network coding: for instance Fragouli et al. [5], illustrate how gains could be obtained compared to routing[1], and protocols have been proposed such as the one from Park et al. [6]. In addition, the literature about network coding, gives *methods* to determine optimal network coding parameters in both wired and wireless networks (see Lun et al. [8] or Wu et al. [7]).

However, while these methods may be used to compute the optimal parameters for network coding in polynomial time, this computation may still be prohibitive for larger networks. Moreover, they do not directly yield insight on the performance of network coding (how much to expect), and how it compares to routing.

In this article, we address these two issues:

- We propose simpler and novel methods to perform network coding based on the knowledge of the topology. They associate connected dominating sets and network coding.
- We prove some results related the performance of these network coding methods.

Our main contribution of this paper is a formal proof of one central aspect of performance of the proposed algorithms. However other main results of this paper might be, arguably, the corollaries deduced from the proven performance: near optimality of the methods in the core of the network (hence of network coding), and comparison between routing and these network coding methods in particular. In general, we also indicate how network coding is expected to outperform routing.

[1] following a common convention, we denote "routing" as "not using network coding", i.e. storing the packets and forwarding them unmodified

The rest of this paper is organized as follows: section 2 provides background material; section 3 describes the methods for network coding, and main performance results; section 4 discusses energy-efficiency; section 5 gives formal proofs of the performance results; and section 6 concludes.

2 Background

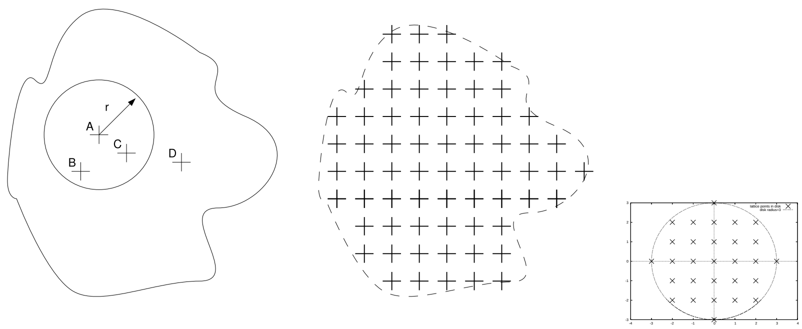
2.1 Problem Statement

In this article, we study the problem of broadcasting from one source in a network to several sources.

2.1.1 Wireless Network Model

Our assumptions are consistent with some commonly found in the broadcast or connected dominating set literature (see for instance, those in [9]).

We consider multi-hop wireless networks with a number of nodes, without mobility. The primary model for the wireless networks that are considered, is the *unit disk graph* model [10] , where two nodes are neighbors whenever their distance is lower than a fixed radio range; see Fig. 1(a) for the principle of unit disk graphs.



(a) Unit disk graph (b) Lattice Network (c) Unit disk range R for lattice

Figure 1: Network Models

Precisely, the wireless networks considered will be:

- Random unit disk graphs with nodes uniformly distributed (Fig. 1(a)) of the plane (or in fact, of Euclidean space $\mathbb{R}^n$)
- Unit disk graphs with nodes organized on a lattice (Fig. 1(b)).

We also assume an ideal wireless model. More precisely, wireless transmissions without loss, collisions or interferences. We assume that each node of the network is operating well below its maximum transmission capacity. Additionally, the network is a packet network and source is assumed to have identically sized packets (so that network coding may be used).

2.1.2 Energy-Efficiency and Rate Selection

We assume one source is present, with an infinite number of packets to transmit.

The goal is to operate in an energy-efficient way, which be formulated as follows. Consider a network at a given time (t), and then consider the number of source packets that have been successfully broadcast to the entire network ($N_p(t)$), and the number of transmissions that have been made by all nodes in the network ($N_t(t)$). The number of transmissions per broadcast is the ratio between the two (that is: $\frac{N_p(t)}{N_t(t)}$). Energy-efficiency corresponds to minimizing this quantity when the time converge towards infinity.

This formulation is equivalent to the classical definition of efficiency of CDS algorithms, where it is the percentage of nodes in the CDS. However with network coding several packets are weaved together, so one cannot apply the direct approach of counting the number of times that a packet has been repeated (i.e. the number of nodes in a CDS).

In the remaining of the article, we will assume that every node has an fixed (average) retransmission rate. This defines the *rate selection.* The metric for evaluating energy-efficiency is the number of transmissions per broadcast. For one source, we count:

- the number of retransmissions from every node, per unit time, directly given by selected rate.
- the number of packets successfully broadcasted from the source to the entire network per unit time;

By dividing the number of retransmissions by the number of packets successfully broadcasted, the metric for the cost per broadcast is obtained. It is denoted E_{cost}.

$$E_{\text{cost}} \triangleq \frac{\text{total transmission rate of all nodes}}{\text{broadcast source rate}} \tag{1}$$

Although it may seem limiting to exclusively consider average rate, it is not, as described in the following section 2.1.3: any network coding method may be converted into a method, as least as efficient, using *random linear coding* with fixed node rates.

2.1.3 Theoretic Grounds for Rate Selection

Several far-reaching results from network coding theory permit to reformulate the problem of energy efficient multicast of a single source, and they may be found for instance in the recent synthesis of Lun et al. [11]. They can be informally described as follows:

- consider *any* network coding method (deterministic, opportunistic, random, ...) on a long duration.
- compute the average rate of the node (packets transmitted per unit time).

Then essentially:

- only the *average rate* of each node has to be considered, to have an lower bound of the cost of the network coding method.
- one will achieve asymptotically this lower bound using the simple method of *random linear coding* of Ho et al. [12], where additionally each node retransmits with a rate equal to the computed average rate.

These results are asymptotic[2] [11]. For energy-efficiency, it follows that the only relevant characterization of any network coding method, is its average rates. Hence the only issue is to choose a set the rates of the nodes that will yield good performance: a *rate selection*.

Once the rate selection is decided, the maximal performance may be computed by purely graph-theoretic methods. Precisely, given any rate selection, one can compute an *maximal achievable broadcast rate* for the source (section 2.3). Essentially, the source may arbitrarily approach this rate and at the same time successfully broadcast all its packets *in the long run*. Hence this is the rate to use in the previous equation (1) as "broadcast source rate".

2.2 Notations

We will use the following general notation in the rest of the article, also illustrated on Fig. 2:

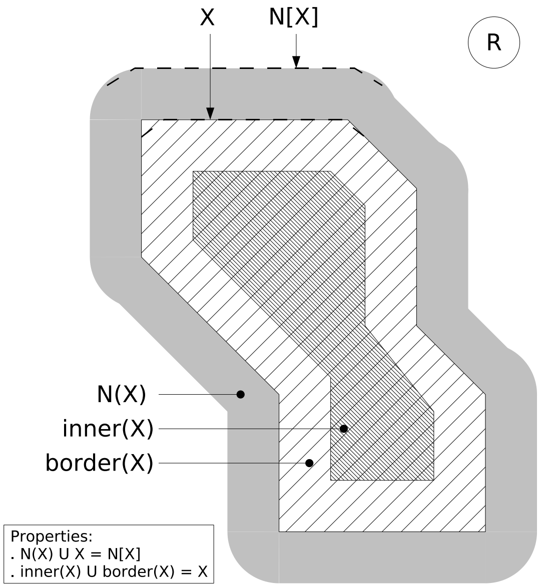


Figure 2: Notations

- Radio range: r (see Fig. 1(a), Fig. 1(c))
- Set of nodes: $\mathcal{V}$
- Rate: the node v retransmits packets with a rate C_v

Some of the notation is more specifically targeted to network of nodes organized on a lattice. Assume that $\mathcal{V}$ is included in a larger set $\widehat{\mathcal{V}}$ (for a lattice, $\mathcal{V} \subset \widehat{\mathcal{V}} = \mathbb{Z}^n$). We use the following notations for concepts related to neighborhood:

- $\mathcal{N}(X)$: open set of neighbors of $X \in \mathcal{V}$; $\mathcal{N}(X) \subset \widehat{\mathcal{V}}$
- $\mathcal{N}[X]$: closed set of neighbors of $X \in V$, that is nodes and their neighbors $\mathcal{N}[X] \triangleq \mathcal{N}(X) \cup X$

[2]as time converge towards infinity, and this is valid for the performance in terms of energy efficiency and not, for instance, decoding delay, CPU cost or other metrics

- inner(X) : the nodes that have only neighbors in X, that is: inner$(X) \triangleq \{x \mid x \in X \text{ and } \mathcal{N}(x) \subset X\}$
- border(X) : nodes that are on the border of X, that is border$(X) \triangleq X \setminus \text{inner}(X)$
- $|X|$: the number of points in the set X when it is finite

For a lattice, in the unit disk model, the set of neighbors of a node is the same as neighborhood of one origin node, represented on Fig. 1(c), with a translation. And we denote:

- R: the (closed) set of neighbors of the origin node
- M_R: the number of neighbors of the origin node, $M_R \triangleq |R| - 1$
- $\mathcal{L}$: the integer lattice, $\mathcal{L} \triangleq \mathbb{Z}^n$ for n integer > 2

We also define notion of *dominating set* as: a dominating set X of a set Y is a subset of Y, such as $Y \subset \mathcal{N}(X)$.

A *connected dominating set* X of a set Y is a dominating set X where the subgraph $X \subset \mathcal{V}$ is connected.

We also introduce two requirements on the network:

- $\mathcal{V}$ is a connected network
- R is a symmetric set (if $x \in R$ then $-x \in R$), a requirement met when R is a unit disk neighborhood.

2.3 Network Coding: Maximum Achievable Broadcast Rate of the Source

In the network coding literature, several results are for multicast, and, apply to the topic of this article, broadcast, since it is special case of multicast. A central result for network coding in wireless networks gives the maximum achievable multicast rate for a single source. It is the rate limit for the source, which ensures that every destination may decode.

The capacity is given by the min-cut from the source to each individual destination of the network, viewed as a hypergraph for wireless networks [8].

Let us consider the source s, and one of the multicast destinations $t \in \mathcal{V}$. The definition of an *s-t cut* is: a partition of the set of nodes V in two sets S, T such as $s \in S$ and $t \in T$. Let $Q(s,t)$ be the set of such *s-t cuts*: $(S,T) \in Q(s,t)$.

We denote ΔS, the set of nodes of S that are neighbors of at least one node of T; the *capacity of the cut* $C(S)$ is defined as the maximum rate between the nodes in S and the nodes in T:

$$\Delta S \triangleq \{v \in S : \mathcal{N}(v) \cap T \neq \emptyset\} \quad \text{and} \quad C(S) \triangleq \sum_{v \in \Delta S} C_v \tag{2}$$

In other terms, the idea is to cut the network into two parts, and check the total rate transmitted from nodes in the part including the source, to nodes of the other part.

The *min-cut* between s and t is the cut of $Q(s,t)$ with the minimum capacity. Let us denote $C_{\min}(s,t)$ as its capacity. From [8], the maximum achievable source rate is given by the minimum of capacity of the min-cut of every destination, $C_{\min}(s)$, with:

$$C_{\min}(s,t) \triangleq \min_{(S,T) \in Q(s,t)} C(S) \quad \text{and} \quad C_{\min}(s) \triangleq \min_{t \in \mathcal{V} \setminus \{s\}} C_{\min}(s,t) \tag{3}$$

2.4 Network Coding: Optimal Rate Selection

Given an instance of the network, one may solve the linear program proposed by [7,8], and find an optimal set of rates (in polynomial time) and the maximum broadcast rate at the same time.

The approach has some scalability issues, as the number of variables N_v is greater than MN^2 (where M is the average number of neighbors, N the number of nodes in the network), and the known worst case complexity for linear programming is $O(N_v^{3.5})$ (as shown by Karmarkar [14]). In practice, both computation time and memory are an issue.

2.5 Related Work

Some results exist about the expected value of the maximal broadcast rate of the source on some classes of wireline networks (with links between pairs of nodes): for instance Ramamoorthy et al. [15] explored the multicast capacity of networks where a source which is two hop from the destinations, through a one network of relay nodes ; Aly et al. [16] studied the some classes of networks in the plane. From their results [15, 16], one intuition is that most nodes have similar neighborhood, hence the performance, when setting an identical rate for each node, deserves to be explored.

For unit disk graphs, and when every node is a source, Fragouli, Widmer and Le Boudec [5] have shown the version of our Th. 3 in the simple case of the torus lattice where nodes have 4 neighbors. Their additional theoretic arguments offer pessimistic guarantees of proper functioning with network coding when rates, hence costs, are higher by a factor of 3 compared to the ones in Th. 3 and Th. 4. Hence they are not sufficient to tightly compare network coding and routing ; and indeed, their results of heuristics for general case indicated good performance with lower rates.

In previous work [20] (extended version: [21]), the authors have established results in the case of a square, where all nodes which are near the edge of the network have higher rate. This was sufficient to prove that network coding is asymptotically locally optimal (see section 4.1.1) when the size of the network converges towards infinity. However, in practice, when the size of the network is fixed, such an approach is inefficient: indeed, in random unit disk graphs, efficiency would converge to zero when density increase. The work presented here is deeper as it uses a more elaborate construction, CDS, to handle appropriatly holes, requiring a new proof. In addition, for random unit disk graphs, we handle the cases where the density is not uniform. Some details of the presented work are also in [22].

3 Our Approach: RAUDS, MARAUDS

3.1 Overview

As mentioned in section 2.1.3, the search for an efficient network coding method is reduced to the choice of a rate selection: deciding the average transmission rate of each node. In this article, we propose two rate selection methods: RAUDS and MARAUDS.

They are derived from one logical argument presented in section 3.2: it starts by considering a network on a lattice (as on Fig. 1(b)), and by considering the assignment the same rate on every node in the hypothetical case where energy-efficiency is perfect in all places.

Naturally, as achieving perfect energy-efficiency is not generally possible with a simple reasoning, the basis of the presented algorithm is the introduction of connected dominating sets to compensate for the problematic places.

Precisely, the CDS are used in three ways: to compensate for the effect of the borders (nonexistent neighbors of nodes in the border); to compensate for occasional lack of neighbors (holes); and to adjust for non-uniform density of the network.

The rate selection methods themselves are presented and formally described in section 3.3 (RAUDS) and section 3.5 (MARAUDS). RAUDS is for networks organized as a lattice (as on Fig. 1(b) and Fig. 3(c)), whereas MARAUDS is an extension for general networks (mapping nodes to a virtual lattice as on Fig. 3(b), and then using RAUDS). They are, in fact, a *family* of methods: rather than being entirely defined, they leave some freedom in the choice of the CDS algorithms and some parameters, and instead they specify conditions that the computed CDS must respect[3].

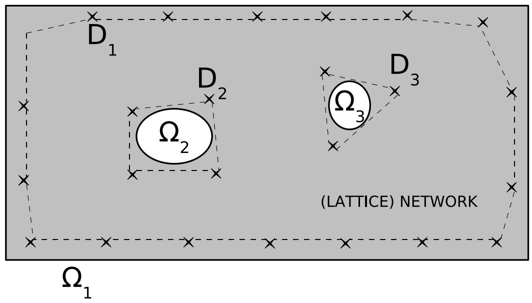
RAUDS starts by setting the source rate as with the logic of section 3.2. For an arbitrary rate selection, the maximum achievable broadcast rate is generally an unknown, which can be computed by graph-theoretic methods, described in section 2.3. However, importantly, RAUDS is such that the conditions posed on the CDS, in fact, can be proved to guarantee that the maximum achievable source rate is the one set by the logic of section 3.2. The results about maximum source achievable rate, for both methods, RAUDS and MARAUDS are summarized in section 3.6. The formal proof itself might be a cornerstone of this article and is given in section 5.

Up to this point, the energy-efficiency of the methods has not yet been addressed. As described in section 2.1.2, it can be computed from both the rates (of the selection) that are directly given by RAUDS and MARAUDS, and the maximum broadcast rate which was proven.

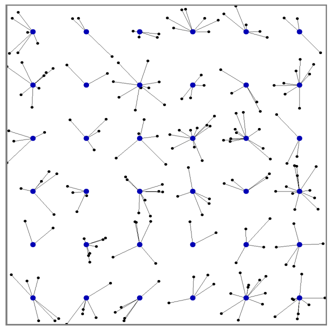
However, the results are slightly deeper: remember that section 3.2 starts with assumption of perfect energy-efficiency in all places; that RAUDS and MARAUDS in section 3.3 and section 3.5 are essentially "correcting" issues in some problematic places; and that section 5 formally proves that the source can still send at the same rate hypothesized in section 3.2. From this, the intuition is that the rate selection could be *almost* perfectly energy-efficient when the problematic places are limited. This issue, of efficiency and closeness to optimality of RAUDS and MARAUDS, is discussed in section 4.

Finally, by noting that RAUDS and MARAUDS are some specific of network coding methods (hence a lower bound of what the network coding might achieve), they shed light on the comparison of the performance of network coding versus the performance of routing, as shown in the same section.

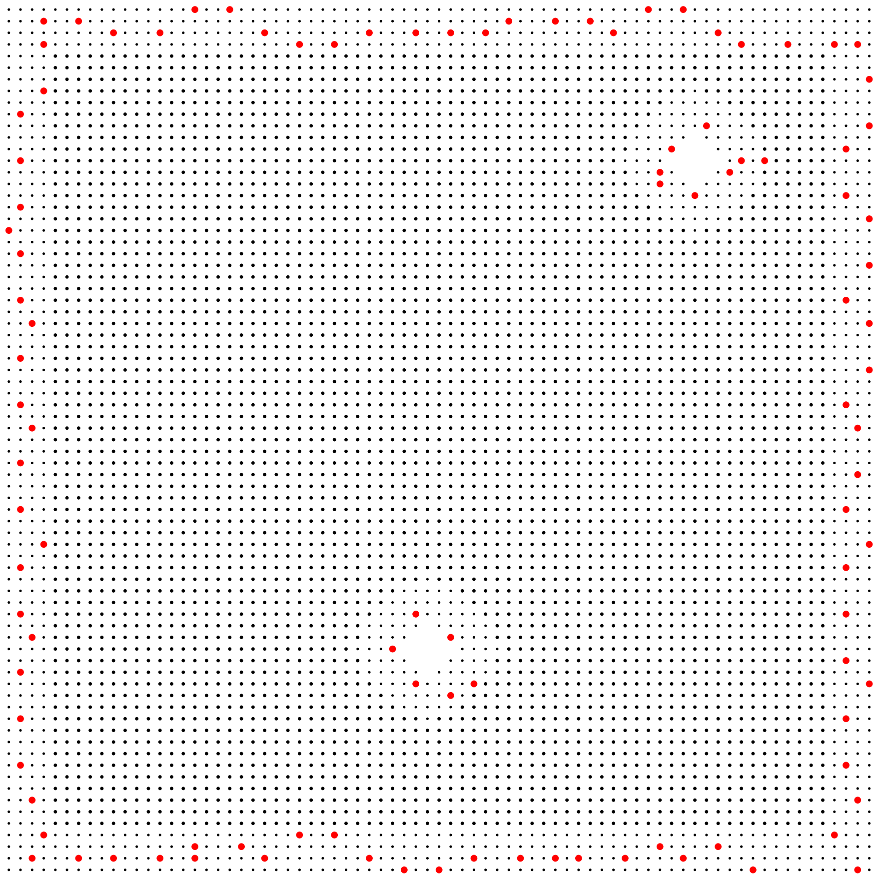
[3]Conversely, it is also shown how any CDS algorithm may be used to in such a way, possibly not optimal, that will respect the conditions.



(a) Principles of RAUDS



(b) Extension of RAUDS to random networks: map to a virtual lattice (MARAUDS)



(c) Example of rate selection with RAUDS - in this case it outperforms any method not using coding

Figure 3: Network Coding with Connected Dominating Sets: RAUDS and MA-RAUDS

3.2 Logic for Rate Selection

Let us start with a definition: in (linear) network coding terms, packet is *innovative* if it will be ultimately decoded into a source packet (otherwise it is not needed nor useful at all).

Now assume a network where the nodes, would be homogeneously distributed, with similar number of neighbors, M. A first approach would be to select the an identical rate for all nodes and improve this rate selection as follows:

1. Assume that every node has an identical retransmission rate, arbitrarily, 1 (e.g. one packet per unit time)
2. Then every node with M neighbors can receive M coded packets per unit time. Assume that nearly all of them innovative
3. Then the source should inject at least M packets per unit time
4. An issue is the nodes of near the border, because they have less neighbors, and in general other nodes with less neighbors, therefore some adjustment is required.

A key step is the adjustment of the rate of the source from 1 to M which allows noticeably higher broadcast rate for the same transmission cost.

3.3 Rate Selection: RAUDS

The rate selection RAUDS is defined initially for nodes organized on a lattice as on Fig. 1(b); section 3.5 details an extension for general networks.

Consider the integer lattice $\mathcal{L} = \mathbb{Z}^n$ and a network $\mathcal{V}$ of nodes which is a subset $\mathcal{V} \subset \mathcal{L}$, such as the one on Fig. 3(c).

The rate selection proceeds in three steps described in the following sections: partition the outside of the network (section 3.3.1), cover problematic nodes with a connected dominating set (section 3.3.2) and collect rate adjustments (section 3.3.3).

The Fig. 3(a) summarizes the outcome of RAUDS and the notation used later, although it should be understood to apply to nodes organized as a lattice as exactly as on Fig. 3(c).

3.3.1 Partition the Outside of the Network

The part of lattice outside of the network $\mathcal{L} \setminus \mathcal{V}$ is partitioned into disjoint sets $(\Omega_i)_{i=1,\ldots,n_\Omega}$, with: Ω_1 is an infinite set and Ω_i for $i \geq 2$ is finite (again as Fig. 3(c)); and also two different sets Ω_i and Ω_j $(i \neq j)$ are disconnected, that is not in range of each other.

At least one such a partition always exists, namely the trivial partition (Ω'_1) with only one set: $n_{\Omega'} = 1$, $\Omega'_1 = \mathcal{L} \setminus \mathcal{V}$

The intent behind using several Ω_i, instead of the previous trivial partition, is to allow each of them match different smaller “holes” in the network and improve efficiency.

3.3.2 Cover Problematic Nodes

The nodes near the Ω_i are covered with connected dominating sets. From the logic in section 3.2. the idea is that such nodes, near the holes Ω_i, will not receive sufficient total rate. The intent is that nodes of the CDS will compensate for this.

This is done as follows: every part of $\mathcal{V}$ in the neighborhood of Ω_i for $i = 1, \ldots, n_\Omega$ is covered by a connected dominating set of nodes D_i, with:
• Each node $u \in D_i$ has a rate adjustment equal to $C^{(D_i)}$
• Each node v neighboring Ω_i (that is $v \in \mathcal{V} \cap \mathcal{N}(\Omega_i)$), must be neighbor of a node of the CDS D_i. *Alternatively, one can relax the property of section 3.3 that the two different sets Ω_i and Ω_j ($i \neq j$) are disconnected, if the following condition is used instead: Each point w neighboring Ω_i (that is $w \in \mathcal{N}(\Omega_i)$), must be neighbor of a node of the CDS D_i, whether it is in $\mathcal{V}$ or in another Ω_j*

The rate adjustment on the CDS depends on the number of missing nodes in the hole Ω_i:
• For Ω_1: $C^{(D_1)} = M_R$
• For $i \geq 1$ and Ω_i: $C^{(D_i)} = |\Omega_i|$ where $|\Omega_i|$ is the size of the hole (it is further capped in next section).

3.3.3 Collect Rate Adjustments

The rate of each node is based on the initial rate 1 with the sum of adjustments of the each CDS in which it is included. Moreover if the sum results into a rate greater than M_R, then it is limited to M_R. Formally:

$$C_v = \min(C_v^*, M_R) \text{ where } C_v^* = 1 + \sum_{i|v \in D_i} C^{(D_i)} \tag{4}$$

For the source s: it has the rate $C_s = M_R$

3.4 Realizations of Rate Selections RAUDS

The previous requirements for RAUDS actually defines a family of rate selections and where degrees of freedom are choice of the (Ω_i), and (D_i) not specified here, allowing to select freely CDS algorithms. We show in section 3.4.1, that it is always possible to realize an (inefficient) rate selection that respect the previous requirements and in section section 3.4.2, we detail how it is possible in general to use any CDS algorithm to create a rate selection RAUDS.

3.4.1 A Trivial Realization of RAUDS

The following simple rate selection RAUDS satisfies the constraints:
• Use the trivial partition: $n_{\Omega'} = 1$, $\Omega'_1 = \mathcal{L} \setminus \mathcal{V}$
• Consider $\mathcal{V} \cap \mathcal{N}(\Omega'_1)$, the set of nodes of $\mathcal{V}$ that have less than M_R neighbors, because they are neighboring Ω'_1
• Choose the dominating set D_1^- initially as $D_1^- \triangleq \mathcal{V} \cap \mathcal{N}(\Omega'_1)$
• Connect the different connected components of D_1^- with some additional nodes from $\mathcal{V}$: the set of these nodes and D_1^- together yield a set D_1 that is a CDS of nodes near Ω_1
• Choose the rate adjustment $C^{(D_1)} = M_R$

3.4.2 A Generic Realization of RAUDS

In practice, one could first choose a connected dominating set algorithm, and operate as follows to construct a rate selection RAUDS: for nodes on the border of the network a CDS would be built with nodes with rate M_R; and also for every large "hole" contained in the network. Whereas for small "holes" in the network, the rate could be locally increased on a few dominating neighbors.

Precisely:

• Consider the set of nodes of $\mathcal{V}$ that are near the border, that is, that have less than M_R neighbors.

• Find the connected components of this set of nodes:
$B_1, B_2, \ldots, B_B$

• Each connected component B_j is either neighboring an infinite set outside $\mathcal{V}$, or a finite hole. Renumber the connected components so that B_1 is bordering the infinite set denoted Ω_1, and denote Ω_j the hole that B_j is bordering.

• For each connected component, create a CDS with the selected CDS algorithm, applied to the set of nodes B_j: this yield the CDS D_j.

• Choose the following rate adjustment: $C^{(D_1)} = M_R$ and for $j \geq 1$, $C^{(D_j)} = |\Omega_j|$

The realization of RAUDS satisfies the constraints. A practical example of applying this exact algorithm is represented by the Fig. 3(c), using one CDS algorithm from Guha and Khuller [2].

3.5 Rate Selection MARAUDS for Unit Disk Graphs

Consider now an arbitrary network which is a random unit disk graph (see Fig. 1(a)). Denote $\mathcal{V}_{\text{real}}$ the set of the nodes on the network, and r the radio range.

The rate selection MARAUDS (*Mapping And Rate Adjustment Using Dominating Sets*) derives from RAUDS by considering a virtual dominating lattice and then by applying RAUDS, as described in the following sections.

3.5.1 Construction of a Dominating Lattice

• Choose some fixed lattice spacing ρ with $0 < \rho < \frac{1}{2}r$ and denote the rescaled lattice as $\mathcal{L}_\rho \triangleq \{\rho\mathbf{x} \mid \mathbf{x} \in \mathbb{Z}^n\}$

• Choose a quantity δ verifying $\frac{1}{\sqrt{2}}\rho < \delta < \frac{1}{2}r$

• Select a mapping λ from the nodes of the real network $\mathcal{V}_{\text{real}}$ to the points of the virtual lattice; $\lambda : \mathcal{V}_{\text{real}} \rightarrow \mathcal{L}_\rho$, which verifies for any $x \in \mathcal{V}_{\text{real}}$:

Property 1. $\|\lambda(x) - x\| \leq \delta$

Such a mapping always exists, for instance, the mapping of nodes of $\mathcal{V}_{\text{real}}$ to the closest node of $\mathcal{L}_\rho$

The *dominating lattice*, denoted $\mathcal{V}$, is then defined as the set of nodes of λ to which a node of $\mathcal{V}_{\text{real}}$ is mapped.

3.5.2 Rate Selection with Dominating Lattice

Define R as a ball, the set of points on the lattice within range r' of the origin, with $r' \triangleq r - 2\delta$ (alternatively, R can be also selected as an arbitrary symmetrical subset of this ball).

It is now possible apply the rate selection *RAUDS* to the virtual network $\mathcal{V} \subset \mathcal{L}_\rho$, with the R chosen. Let $C_v^{(\text{lat})}$ be a rate selection such as one described in section 3.3 for the lattice. The rate selection for the initial network $\mathcal{V}_{\text{real}}$ is then:

- For any node $v \in \mathcal{V}_{\text{real}}$, a rate C_v is chosen, respecting the following property:
  the sum of the rates of nodes mapped to the same point τ of the dominating lattice is equal to the rate $C_\tau^{(\text{lat})}$ of that point for the rate selection of the dominating lattice. Formally:
$$\sum_{v \in \mathcal{V}_{\text{real}} | \lambda(v) = \tau} C_v = C_\tau^{(\text{lat})} \tag{5}$$

One straightforward way to ensure (5) is the following: for every point of the dominating lattice $\tau \in \mathcal{V}$, choose one unique $v \in \mathcal{V}_{\text{real}}$ mapped to that point, and assign the rate $C_v = C_\tau^{(\text{lat})}$. Essentially, it means that a small dominating set of nodes is selected, spatially organized as a (near-)lattice. For high density networks, this implies that the number of retransmitting nodes is limited, and also importantly, it provides an efficient way to handle non-uniformity.

3.6 Maximum Source Broadcast Rate of RAUDS and MARAUDS

A central question for the evaluation of the energy-efficiency of RAUDS and MARAUDS, as seen from (1), is the maximum broadcast rate of the source.

The actual maximum source broadcast rate is also an assessment of how well the rate selections approach the hypothetic reasoning of section 3.2. Indeed, by construction, the rate selections aim at ensuring a fixed maximum broadcast rate, equal to M_R. The essence of our main result is the proof that they succeed, in section 5, Th. 3:

Property 2. *Assume a rate selection RAUDS, defined by $(\Omega_i), (D_i), (C_v^{(D_i)})$, verifying (4), then maximum source broadcast rate is greater or equal to M_R. It is exactly M_R when at least one node is not neighbor of the source nor the dominating sets.*

The proof relies on fundamental properties of discrete sets of the Euclidean space (15) used for the proof of the intermediary Th. 5 in section 5.7.

Similarly in section 5.5, Th. 4 proves that:

Property 3. *The maximum broadcast rate of the rate selection MARAUDS of the section 3.5 is $\geq |R| - 1$*

Notice that this is an inequality, and no longer an equality. The reason is that R is not given, but constructed. As a result if the parameters are too coarse (ρ and δ), the proven broadcast rate is lower that the actual broadcast rate. In practice when both $\rho \ll r, \Delta r \ll r$, as intended, the performance is close.

4 Energy-efficiency of RAUDS, MARAUDS and Network Coding

In this section, we discuss the energy-efficiency of RAUDS and MARAUDS: we start with an insight in section 4.1 of why the method RAUDS is locally optimal for energy-efficiency in distant areas from the borders. For (infinitely) dense networks of the plane, this is used in section 4.2.1, to give an informal evaluation of the energy-efficiency MARAUDS derived from the previous insight, and in section 4.2.2 we summarize the results formally proved in the case of the square network.

In the last sections, section 4.3 and section 4.4, we use the fact that routing cannot be locally optimal, and give tight bounds for its performance: as a result, we show how MARAUDS or network coding in general are expected to outperform routing.

4.1 Energy-efficiency and Local Optimality of RAUDS

4.1.1 Local Optimality of RAUDS

To have an intuition of the impact of the previous results with respect with energy-efficiency, consider a network organized as a lattice, with a rate selection RAUDS ; such as the one represented on Fig. 3(c) for instance.

Consider one node u in the "*inside*" of the network, that is, a node that would be sufficiently far from the borders, the dominating sets, and the source, so that all its neighbors will have a rate equal to 1. It has M_R neighbors. Denote informally by *interior nodes* such nodes.

Th. 3 (summarized in previous section 3.6) states that the maximum source broadcast rate is M_R. Notice that every rate is an integer, and now if the network also operates synchronously (as defined in [17]), one can apply the results of Jafarisiavoshani et al. [17] (their Th. 1):the interior node u will receive innovative packets from the source at a rate exactly M_R, after a transition phase, provided that linear network coding is done with a sufficiently large field size.

Since in a synchronous operating mode, the node u will receives also exactly M_R packets per unit time (one from each neighbor), it follows that every of them will be innovative after the transition phase, exactly as hypothesized in section 3.2.

Conversely, consider now the perspective of the sender. Consider a node v such that its neighbors are interior nodes[4] like u: then after a transition period, conversely, every of its transmissions will be innovative for all its receivers, its neighbors.

At this point, remark that on the lattice, no node may have M_R neighbors, so no transmission from any node may provide innovative information to more than M_R nodes.

From the perspective of energy-efficiency, the objective is minimize the number of transmissions, so equivalently, one has to globally maximize the number of nodes for which the transmissions are innovative. Then, the direct implication is that the transmissions of the node v are locally optimal from the energy-efficiency point of view.

In this section, for simplicity in the explanation, we relied on results of [17], but the intuition presented is still asymptotically true in general, for any retransmission scheduling, and any field size (see [11]).

[4] in other terms, the node v is at least three-hops away from the dominating sets and also the source

4.1.2 Simple Bound For Energy-Efficiency

Rewritten more formally, the reasoning in the previous section 4.1 gives a simple bound on the minimum number of transmissions per broadcast E_{cost}: assume that every node has at most $M_{\max}$ neighbors; one single transmission is received by and useful to at most $M_{\max}$ nodes. Hence, in order to broadcast one packet to all N nodes, at least $E_{\text{bound}} = \frac{N}{M_{\max}}$ transmissions are necessary. One metric for energy-efficiency can be equivalently written as the ratio to this bound: $\text{Eff}_{\text{bound}} \triangleq \frac{E_{\text{bound}}}{E_{\text{cost}}}$. Note that in general, $\text{Eff}_{\text{bound}} = 1$ exactly may not be achieved by any method.

4.2 Energy-efficiency of MARAUDS for Dense Networks of the Plane

4.2.1 Insight on the Performance of MARAUDS

We focus on networks which are contained into a predefined region of the plane, and we are interested in the energy-efficiency of the method MARAUDS when the density increases whereas the range stays fixed. As the density increases, any point of the space will be arbitrarily close to a node of the network, hence geometric reasonings are used.

Consider networks contained in regions such as the one represented on Fig. 2: the networks themselves would inside the region denoted X. The part denoted inner(X) is the region where the nodes would have a sufficient number of neighbors as the density increases: it is the part where, asymptotically, MARAUDS would be locally optimal. The part border(X) is the region where nodes need to be covered by CDS. The idea is that the border(X) can be covered by nodes standing near the border as represented on Fig. 4. In that case, we can infor-

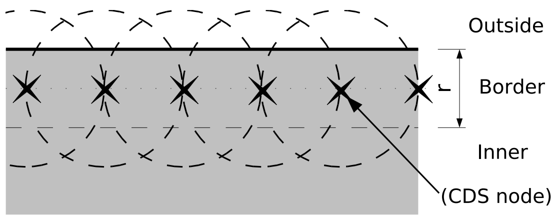


Figure 4: CDS to cover borders in dense networks

mally estimate the number of nodes in the CDS from the perimeter P of the line midway between the edge of X and the edge of inner(X) as: $N_{\text{cds}} \approx \frac{P}{r}$. Assume that the area of the network is A, and that the virtual lattice of MARAUDS has a step ρ. Then $N_{\text{lattice}} \approx \frac{A}{\rho^2}$, $M \approx \frac{\pi r^2}{\rho^2}$, $E_{\text{cost}} \approx \frac{1}{M}(\frac{P}{r} \times M + N \times 1)$. Finally denoting as $\lambda \triangleq \frac{A}{P}$, the ratio between area and perimeter of region of the network, we can estimate the efficiency bound of section 4.1.2:

$$\text{Eff}_{\text{bound}}^{(\text{est.})} = \frac{1}{1 + \frac{\pi r}{\lambda}} \tag{6}$$

Notice that when r is kept constant and the entire region of network is scaled by α, the ratio area/perimeter λ is scaled by the same amount α. Hence $\text{Eff}_{\text{bound}}^{(\text{est.})} \to 1$ when the area of the network grows indefinitely; this is because the area of the border becomes vanishingly small in comparison. This indicates asymptotic optimality of MARAUDS (as for RAUDS). For the case where the network is included in square region, a formal proof of (6) is in next section.

4.2.2 Asymptotic Performance of MARAUDS

In this section, we will establish performance bounds of MARAUDS, for growing density, but fixed network region. Precisely: assume networks contained in the plane, inside a square region $\mathcal{A}$ of width L and with a fixed radio range $r > 0$. Consider a rate selection MARAUDS as in section 3.5, with the following parameters:

- $\rho = \frac{1}{\mu^{\frac{1}{3}}}$ (μ defined later) and $\delta = \rho$
- the nodes of $\mathcal{V}_{\text{real}}$ are mapped to the closest point on the rescaled lattice $\mathcal{L}_\rho$. Denote $\mathcal{V}(\rho) = \lambda(\mathcal{V}_{\text{real}})$.

For an asymptotic result, we consider a sequence of random unit-disk graphs where the nodes are defined by spatial Poisson process with rate μ, with $\mu \to \infty$. We are interested in the bound given in section 4.1.2, and the estimate of section 4.2.1, which is proven as the following theorem:

Theorem 1. *The efficiency of the MARAUDS is such that:* $E[\text{Eff}_{\text{bound}}] \xrightarrow{p} \frac{1}{1+\frac{4\pi r}{L}}$ *for an appropriate rate selection MARAUDS, when the density* $\mu \to \infty$

Proof. See section 5.6 □

For non-uniform networks: consider a sequence of non-uniform networks indexed by $i \in \mathbb{N}$, $i \to \infty$, and assume that the nodes are given by point processes with of density $\mu_i(x)$ for $x \in \mathcal{A}$. We have the following additional theorem:

Theorem 2. *If the densities verify* $\min_{\mathcal{A}} \mu_i \to \infty$*, the efficiency of the MARAUDS is such that:* $E[\text{Eff}_{\text{bound}}] \xrightarrow{p} \frac{1}{1+\frac{4\pi r}{L}}$

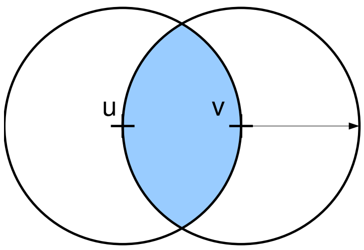
Proof. We can consider the $\mu = \min_{x \in \mathcal{A}} \mu_i(x)$ in the proof of Th. 1: then the proof also applies. What happens is that, with MARAUDS, every region of the plane will contain the same density of nodes of the dominating lattice $\mathcal{V}$. □

4.3 Comparison of the Energy-Efficiency of Network Coding and Routing

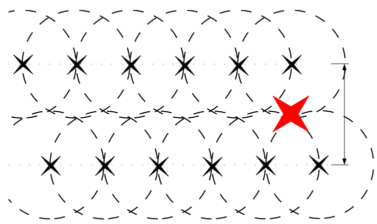
In section 4.1, we described the local optimality of RAUDS, MARAUDS in the parts of the networks that are far (three-hop or more) from the CDS and the source.

One question is: without network coding, can the broadcast also be locally energy-efficient (on the lattice and in general)? One answer is negative, from an argument of Fragouli et al. [5]: consider any broadcast with a store-and-forward, non-coding method (including, but not limited to, the use of CDS for broadcast), and consider the sets of nodes which have (re-)transmitted the packet. Apart from the source, any retransmitting node v, will have received the packet from another node u. When v retransmits the packet, its transmission will be received again by u itself, and additionally any common neighbors of u and v: for those nodes the transmission will be redundant (see Fig. 5(a)). Hence, apart from the source the retransmissions are never locally optimal.

If we consider dense networks in the Euclidean plane, and use areas to evaluate the bound derived from the previous argument: in a unit disk graph with range r, two neighbors share a neighborhood area at least equal to $(\frac{2\pi}{3} - \frac{\sqrt{3}}{2})r^2$ (Fig. 5(a)), hence a bound for the transmission-level efficiency of broadcast in a dense, uniformly distributed graph, is:
$\mathrm{Eff}^{(\mathrm{routing})}_{\mathrm{upper-bound}} = \frac{E^{(\mathrm{routing})}_{\mathrm{bound}}}{E^{(\mathrm{routing})}_{\mathrm{cost}}} \leq \frac{2\pi+3\sqrt{3}}{6\pi} \approx 0.609$ (see also [5]).



(a) Inefficiency of routing



(b) CDS cover of the plane

Figure 5: Routing: bound and cover

On the other hand if the area tends towards infinity, it is at least possible to cover the whole space with alternating rows of aligned nodes such as the border CDS of Fig. 4[5]. This yields an lower bound of the possible efficiency of routing: $\mathrm{Eff}^{(\mathrm{routing})}_{\mathrm{lower-bound}} = \frac{2+\sqrt{3}}{2\pi} = 0.594\ldots$

Hence for infinitely large and dense networks, and for single source broadcast, we have the following bounds for the gain of network coding over non-coding:

$$1.642 \approx \frac{6\pi}{2\pi + 3\sqrt{3}} \leq \text{coding gain} \leq \frac{2\pi}{2+\sqrt{3}} \approx 1.684 \quad (7)$$

4.4 Comparison of MARAUDS and Routing

In the previous section, the comparison of network coding and routing was possible when the area of the network would grow indefinitely, because at the same time, MARAUDS would become asymptotically optimal.

Another question, would be to estimate if gains over routing would be expected for a given area of the network (but the density still increases). Our results cannot answer this question, since the efficiency of the optimal rate selection for network coding are not known, however, some insight is possible for the rate selection MARAUDS, from (6) and Th. 1. Comparing Th. 1 and the upper bound of efficiency for routing $\mathrm{Eff}^{(\mathrm{routing})}_{\mathrm{upper-bound}}$ of previous section, it follows that MARAUDS would be expected to be advantageous over routing at least for dense square networks where:

$$\frac{r}{L} \leq \frac{3}{4\pi+6\sqrt{3}} - \frac{1}{4\pi} \approx 0.0510\ldots \text{ that is where: } r \leq \frac{1}{20}L$$

[5]their spacing is $r(1+\frac{\sqrt{3}}{2})$, and the cost of additional connecting nodes such as the colored one in Fig. 5(a) is $O(\frac{1}{\lambda})$

4.5 Sample Rate Selection with RAUDS

The Fig. 3(c) represents one example of rate selection RAUDS. The larger (colored) points represent nodes on the CDS with higher rate, whereas other nodes have rate 1. The points with slightly smaller size are nodes on the borders that are covered by 3 CDS. The source may be placed at any node.
• There are $N = 5599$ nodes, range is $r = 4$ and $M_R = 48$
• Two neighbors have at most 30 non-common neighbors, and then the lower bound on the size of a CDS is 187
• The cost per broadcast of the RAUDS rate selection is $E_{\text{cost}}^{(\text{RAUDS})} = 186.1875$, hence outperforms any method not using coding
• For reference the cost per broadcast of two CDS methods applied to the whole area are:

- Guha and Khuller CDS [2]: $E_{\text{cost}}^{(\text{GK})} = 292$
- Dai and Wu CDS [4]: $E_{\text{cost}}^{(\text{DW})} = 349$

This example is illustrative because it is one for which RAUDS will outperform any method using routing. The gap with the efficiency of the upper bound of routing and RAUDS is rather small: one reason is that the parameters $\frac{r}{L}$ are slightly lower than those given in section 4.4 (here $\frac{r}{L} = 18.75$), beyond which there would be an expectation for RAUDS or MARAUDS to gain advantage over routing. Finally, on this example, note that some well-known CDS algorithms are far from achieving the upper bound (by 50 and 80 %).

4.6 Extensions

• Scaling the rates: the rates may be scaled by the same amount ; the energy efficiency stays identical
• Several broadcast sources: this is easily handled using the construction of [1] of a common virtual source.
• Euclidean spaces: results of in section 3.6 are valid for any Euclidean space (any dimension), and any symmetrical set R.
• RAUDS essentially is a construction of a connected dominating set to compensate for “holes” in a network where the capacity of the cut was otherwise known (without holes). These are easily generalized to other topologies than the Euclidian integer lattice. As an example, the proofs may be immediatly applied to topologies where Th. 5 is verified.

5 Proofs for RAUDS and MARAUDS

5.1 Overview of the Proofs For Maximum Broadcast Rate

In this section 5, a proof for the maximum broadcast rate of the source with the rate selection RAUDS is given. This is done by considering the capacity of the min-cut of the source, and by proving that it is M_R. The min-cut can be determined from the capacity of the cut (S, T) of the network.

The main idea is to use properties of discrete geometry on the Euclidean lattice $\mathcal{L}$ provided by Th. 5.

But in order to do so, a proper partition of the full lattice $\mathcal{L}$ is needed, whereas the $(S, T) \in Q(s, t)$ considered for the computation of the min-cut are only partitions of the subset $\mathcal{V}$ of $\mathcal{L}$. The idea, developed in section 5.2, is to construct an extended partition of $\mathcal{L}$ from any partition (S, T) of $\mathcal{V}$.

Then one lemma (lemma 2) essentially proves that for the extended partition the capacity of the corresponding cut (considering $\mathcal{L}$ as the network) would be sufficient, owning to the large number of neighbors.

When the initial cut (S, T) is considered again, compared to the extended partition, some of the nodes of the extended partition will be missing; hence the capacity of the cut would be lower. The central idea of lemma 3 and its proof in section 5.3, is to ensure that the dominating sets chosen by the rate selection compensate for the missing nodes from the extended partition. The theorem then follows.

5.2 Extended partition S^*, T^* of a cut S, T

Consider $t \in \mathcal{V}$ and one arbitrary s-t-cut $(S, T) \in Q(s, t)$.

We expand S, T, partition of $\mathcal{V}$, into a *extended partition* S^*, T^* of $\mathcal{L}$ in a manner defined here.

Recall that the $\cup_{i=1,\ldots,n_\Omega} \Omega_i$ are a partition of the space outside $\mathcal{V}$. The idea is that, by deciding how to assign them either to S^* or to T^*, a partition of $\mathcal{L}$ could be constructed. We proceed as follows:

- if Ω_i has neighbors only in T, then it will be in T^*,
- otherwise if the associated dominating set D_i is fully included in T, it will be in T^* as well,
- otherwise it will be in S^*.

Formally, the indices $i \in \{1, \ldots, n_\Omega\}$ of the Ω_i corresponding of each of the three previous cases are grouped accordingly into separate sets $I^{(T)}_{\text{incl}}$, $I^{(T)}_{\text{dom}}$ and $I^{(S)}$:

- $I^{(T)}_{\text{incl}} = \{i | \mathcal{N}(\Omega_i) \subset T \text{ with } i \in \{1, \ldots, n_\Omega\}\}$
- $I^{(T)}_{\text{dom}} = \{i | D_i \subset T \text{ with } i \in \{1, \ldots, n_\Omega\} \setminus I^{(T)}_{\text{incl}}\}$
- $I^{(S)} = \{1, \ldots, n_\Omega\} \setminus (I^{(T)}_{\text{incl}} \cup I^{(T)}_{\text{dom}})$

Then the partition S^*/T^* of the lattice $\mathcal{L}$, previously described, can be specified formally as:

$$T^* \triangleq T \cup \Big(\bigcup_{i \in I^{(T)}_{\text{incl}} \cup I^{(T)}_{\text{dom}}} \Omega_i \Big) \text{ and } S^* \triangleq S \cup \Big(\bigcup_{i \in I^{(S)}} \Omega_i \Big) \tag{8}$$

By construction of the extended partition we have the following property:

Lemma 1. *Consider an extended partition S^*, T^* of a cut S, T, as defined previously. Then, any node of S^* which has a neighbor in T^* must have a neighbor in T. Formally, this property can equivalently be written as:*

$$\text{border}(S^*) \subset \mathcal{N}(T)$$

Proof. Consider any such $v \in \text{border}(S^*)$. By definition, it has at least one neighbor in T^*, and consider such a neighbor u. By definition of T^*, there could be three possibilities:

- $u \in T$: in that case, indeed, v is neighbor of a node in T
- $u \in \Omega_i$ with $i \in I^{(T)}_{\text{incl}}$: this case is not actually possible since by definition of $I^{(T)}_{\text{incl}}$, $\mathcal{N}(\Omega_i) \subset S$, which implies $\mathcal{N}(u) \subset T$, and hence u cannot be neighbor of $v \notin T$

- $u \in \Omega_i$ with $i \in I^{(T)}_{\text{dom}}$: in this case, v is neighbor of Ω_i; by construction of the dominating set D_i, the nodes from D_i are covering all the neighbors of Ω_i, hence v must be neighbor of some $w \in D_i$. Since for $i \in I^{(T)}_{\text{dom}}$, by definition $D_i \subset T$, it follows that $w \in T$ and therefore v is neighbor of T.

In all cases, we have proven that v has a neighbor in T, hence the lemma. □

5.3 Relationship between Capacities of Cut and Extended Partition

The general idea of the following lemma, is to show that there are two cases: either the nodes of S^* which are neighbors of nodes of T are in sufficient number (and we will later consider their contribution to the min-cut), or the source $s \in S$ is a neighbor of T. Namely:

Lemma 2. *Consider an extended partition S^*, T^* of a cut S, T, as defined in section 5.2. Then we have one of the following properties:*

- *either s is neighbor of T*
- *or $|\text{border}(S^*)| \geq M_R$*

Proof. There are two cases: either S^* is finite or T^* is finite depending on which of them includes Ω_1[6]:

• If S^* is finite, then by applying Th. 5 with $X = S^*, Y = T^*$, there are two possible sub-cases from (12):

- in the first sub-case, $\text{border}(S^*) = S^*$. Then combined with lemma 1, we have: $S^* \subset \mathcal{N}(T)$ Applying this property to the source $s \in S \subset S^*$, we have: s is direct neighbor of at least a node in T.
- otherwise, $|\text{border}(S^*)| \geq M_R$

• If T^* is finite, from Th. 5 with $X = T^*, Y = S^*$, we have from (13): $|\text{border}(S^*)| \geq M_R$

Hence the lemma. □

We are now able to prove the main result of the capacity of the s-t cut S, T: if S^* was equal to S, this would be sufficient to establish the desired property on the capacity on the cut, because of the preceding lemma, as there would be more than M_R nodes of S^* contributing with a rate ≥ 1.

However this is not the case in general because of the holes $\Omega_j \subset S^*$. This is where the dominating sets are used to compensate for the existence of the Ω_j. The following lemma proves the actual result:

Lemma 3. *Consider an s-t-cut S, T, with the rate selection RAUDS. The capacity of the cut $C(S)$ verifies:*

$$C(S) \geq M_R$$

Proof. Create an extended partition S^*/T^* as in section 5.2.

From lemma 2, either s is a neighbor of T, or $|\text{border}(S^*)| \geq M_R$. In the first case, s contributes directly to the capacity of the cut $C(S)$ with a rate M_R, hence it does not constraint the min-cut. Likewise, for any considered node

[6] Recall that by definition Ω_1 is the only infinite subset of $\mathcal{L}$ among: S, T, Ω_i for $i = 1, \dots n_\Omega$

in the following, that has a rate adjustment greater than M_R, so that it was limited to to M_R - any such node will contribute directly to the cut $C(S)$ with a rate M_R, hence would not constraint the min-cut, and the desired result is obtained. Hence such cases will be ignored.

In any case, we can focus on the second case, where:

$$|\text{border}(S^*)| \geq M_R \tag{9}$$

In the following, we will prove that $C(S) \geq |\text{border}(S^*)|$.

For any $i \in I^{(S)}$, consider ω_i, defined as the part of Ω_i which near T^*: $\omega_i \triangleq \Omega_i \cap \text{border}(S^*)$. We will find a lower bound for the compensation rate C_{ω_i} that the nodes in ω_i are sending, to nodes in T.

Three cases need to be considered:

- either ω_i is empty. Then we choose $C_{\omega_i} = 0$
- either $\omega_i \neq \emptyset$ and the dominating set D_i is fully included in S^*.
- either $\omega_i \neq \emptyset$ and $D_i \not\subset S^*$

In the first case, Ω_i is not neighboring T^*, hence has no impact on the capacity of the cut S/T.

In the second case $D_i \subset S^*$: choose one node $u_i \in \omega_i$. It must be such as $u_i \in \text{border}(S^*)$ and from lemma 1, there exists at least one neighbor node $v \in \mathcal{N}(u_i)$, verifying $v \in T$. By definition of the dominating set D_i, the node v must receive a total extra rate $C_{\omega_i} \geq |\Omega_i|$.

In the third case $D_i \not\subset S^*$: D_i is not fully included in T^* either, because otherwise i would be in $I^{(T)}_{\text{dom}}$, and not $I^{(S)}$. Consider the partition of D_i, $(D_i \cap S^*,\ D_i \cap T^*)$ where none of these sets are empty. By definition of the dominating set D_i, the total extra rate received by the set $D_i \cap T^* \subset T$ must be $C_{\omega_i} \geq |\Omega_i|$.

Now consider the definition of the capacity of the cut $C(S)$. From (2):

$$C(S) = \sum_{v \in \Delta S} C_v$$

with, ΔS are the nodes of S, neighbors of nodes of T from (2): $\Delta S = \{v \in S | \mathcal{N}(v) \cap T \neq \emptyset\}$

Now let us compare ΔS and border(S^*). Every node of S which is a neighbor of a node of T (those in the set ΔS), is also a point of S^* which is a neighbor of a point of T^*, because these are supersets. Hence $\Delta S \subset \text{border}(S^*)$. Conversely, by lemma 1, the part of $S \subset S^*$ neighboring T^* must also be neighbor of nodes in T, thus, $\text{border}(S^*) \cap S \subset \Delta S$. Hence actually $\Delta S = S \cap \text{border}(S^*)$.

Now consider border(S^*). We can rewrite:

$$\begin{aligned}
\text{border}(S^*) &= S^* \cap \text{border}(S^*) = (S \cup \bigcup_{i \in I^{(S)}} \Omega_i) \cap \text{border}(S^*) \\
&= (S \cap \text{border}(S^*)) \cup (\bigcup_{i \in I^{(S)}} \Omega_i \cap \text{border}(S^*)) \\
&= \Delta S \cup (\bigcup_{i \in I^{(S)}} \omega_i)
\end{aligned}$$

$$\text{Therefore :} \qquad |\text{border}(S^*)| = |\Delta S| + \sum_{i \in I^{(S)}} |\omega_i| \tag{10}$$

From the expression of the capacity of the cut:

$$\begin{aligned} C(S) &= \sum_{v\in\Delta S} C_v \overset{(a)}{\geq} |\Delta S| + \sum_{i\in I^{(S)}} C_{\omega_i} \\ &\overset{(b)}{\geq} (|\text{border}(S^*)| - \sum_{i\in I^{(S)}} |\omega_i|) + \sum_{i\in I^{(S)}} C_{\omega_i} \\ &\overset{(c)}{\geq} (|\text{border}(S^*)| - \sum_{i\in I^{(S)}} |\omega_i|) + \sum_{i\in I^{(S)}} |\Omega_i| \\ &\geq |\text{border}(S^*)| + \sum_{i\in I^{(S)}} (|\Omega_i| - |\omega_i|) \\ &\overset{(d)}{\geq} |\text{border}(S^*)| \overset{(e)}{\geq} M_R \end{aligned}$$

(a) comes from considering all nodes in T: as a whole, they receive the base rate 1 from all nodes is ΔS, and in addition, the compensation rates C_{ω_i} described previously.
(b) comes from (10); (c) comes from the established property $C_{\omega_i} \geq |\Omega_i|$; ($d$) from the fact that $\omega_i \subset \Omega_i$; and ($e$) from (9)

This yields the lemma. □

5.4 Maximum Broadcast Rate for RAUDS

The main result is a corollary of the previous result:

Theorem 3. *Assume a rate selection RAUDS, defined by* $(\Omega_i), (D_i), (C_v^{(D_i)})$, *then maximum broadcast rate of the source* $\geq M_R$. *It is exactly* M_R *when at least one node is not neighbor of the source nor the dominating sets.*

Proof. Consider any s-t-cut $(S, T) \in Q(s, t)$: lemma 3 indicates that its capacity verifies $C(S) \geq M_R$. Hence the capacity of the min-cut $C_{\min}(s, t)$ and the maximum broadcast rate $C_{\min}(s)$ verify the same property, which is the first part of the theorem.

If at least one node u is not neighbor of the source nor the dominating sets, then we considering the cut $S_0 = \mathcal{V} \setminus \{u\}, T_0 = \{u\}$. We have $C(S_0) = M_R$ therefore the lower bound is reached; hence the theorem. □

5.5 Maximum Broadcast Rate for MARAUDS

Theorem 4. *The maximum broadcast rate of the rate selection MARAUDS proposed in the section 3.5 is* $\geq |R| - 1$

Proof. For $s \in \mathcal{V}_{\text{real}}, t \in \mathcal{V}_{\text{real}}$, consider a cut $(S, T)$of the graph $\mathcal{V}_{\text{real}}$. Denote $s' = \lambda(s)$ and $t' = \lambda(t)$.

An *induced cut* (S', T') of the $\mathcal{V} \subset \mathcal{L}_\rho$, the dominating lattice, is constructed as follows:

- For any point of the lattice $\tau \in \mathcal{V}$, the rate is $C_\tau^{(\text{lat})}$.
- S' is the set of the points of $\mathcal{V}$ such as only nodes of S are mapped to them:

$$S' \triangleq \{\tau : \lambda^{-1}(\tau) \subset S\} \tag{11}$$

- T' is the set of the rest of points of $\mathcal{V}$.

Now, if $s' \in T'$, this implies that the source s is mapped to the same node s' as at least one other node $v \in T$. This implies that the node is neighbor of the source, hence, the capacity of the cut $C(S)$ is at least the rate of the source, which is sufficient to establish the theorem in this case.

Then the only case that need to be considered is the case where $s' \in S'$. Notice that $t' \in T'$; that all the points of the lattice, to which both nodes from S and T are mapped, these points are in T'. Therefore S', T' is indeed a partition and a $s' - t'$ cut.

By Th. 3, we know that the capacity of the cut S', T' is lower bounded by $M_R = |R| - 1$: $C'(S') \geq |R| - 1$

Consider the expression of the capacity of the cut (S', T'): From (2): $C'(S') = \sum_{v \in \Delta S'} C_v^{(\text{lat})}$
where $\Delta S'$ are the points of S', neighbors of points of T' (from (2)).

Consider two points in $u' \in \Delta S'$ and one associated neighbor $v \in T'$. Now consider any two nodes of $\mathcal{V}_{\text{real}}$ that are mapped to those points: $u \in S$ and $v \in T$, such as $\lambda(u) = u'$ and $\lambda(v) = v$. Such nodes exist by construction.

We have:

$$\begin{aligned} \|u - v\| &\leq \|u - u'\| + \|u' - v'\| + \|v' - v\| \\ &\leq \|u - \lambda(u)\| + \|u' - v'\| + \|v - \lambda(v)\| \\ &\leq 2\delta + \|u' - v'\| \text{ (with Property 1)} \leq r \end{aligned}$$

Hence, u will also be in ΔS as defined in (2) and contributes to the capacity of the cut of $C(S)$. Considering all other nodes of S mapped to the same u', (5) indicates that their total rate is equal to $C'_{u'}$, and they are also neighbors of v. Therefore their total contribution to the capacity of the cut $C(S)$ is identical to the contribution of u' to the cut $C'(S')$.

Applying the same reasoning to all nodes in S', we have: $C(S) \geq |R| - 1$. This establishes the theorem. □

5.6 Proof of Th. 1 in section 4.2

Consider a lattice with the previous rate selection. $\mathcal{V}(\rho)$ is the set of points of the lattice $\mathcal{L}_\rho$ to which at least one point of $\mathcal{V}$ is mapped. When the rate of the spatial Poisson process μ is large enough, this sets corresponds to the set of points of the full $\mathcal{L}_\rho$ inside the square containing the network. We first select ρ to verify this property.

Precisely: let us denote $\mathcal{E}_0$ the event, for one point of the lattice: *"there is no point mapped to it"*. We have $\Pr[\mathcal{E}_0] = e^{-\mu\rho^2}$. If $\mathcal{E}$ is the global event *"at least one point of the lattice has no point mapped"*, an *union bound* on the $\frac{L^2}{\rho^2}$ points of the lattice $\mathcal{L}_\rho$ yields $\Pr[\mathcal{E}] \leq \frac{L^2}{\rho^2} \Pr[\mathcal{E}_0]$, hence:
$\Pr[\mathcal{E}] \leq \exp(-\mu\rho^2 - 2\log\rho + 2\log L)$

By setting $\rho = \frac{1}{\mu^{\frac{1}{3}}}$ for instance, we have the desired property $\Pr[\mathcal{E}] \to 0$ when $\mu \to \infty$.

Now consider the efficiency $\text{Eff}_{\text{bound}}$, which involves E_{cost}, the "transmissions per broadcast.", as defined in section 2.1.2, which in turns requires the maximum broadcast rate, and the total transmission rate T_{cost}

From Th. 4, we have: the maximum broadcast rate is $|R(\rho)|-1$. The number of points in $|R(\rho)|$ is the number of lattice points within a circle of radius fixed around the origin (the "*circle problem*"); it is $|R(\rho)| = \pi(\frac{r}{\rho})^2 + O(\frac{1}{\rho})$ when $\rho \to 0$ ([19] p. 133)

The rate of transmissions T_{cost} is given by the rate of transmissions of nodes in the network, plus the rate of transmissions of the nodes on the dominating set, and the source. When the event $\mathcal{E}$ is verified, there is no "hole" and only the nodes on the border of $\mathcal{V}$ require rate adjustment from their neighbors.

We can construct a dominating set composed of nodes of $\mathcal{L}_\rho$ on four lines parallel to the edges of the square defining the network area (see Fig. 4).

Their space is chosen as $\approx r$, i.e. $r + O(\rho)$, and from elementary geometry, this is always possible if ρ is small enough (from elementary geometry, sufficient condition: $\rho \leq \frac{r}{8}$), and the total number of nodes in the dominating set is $\frac{4L}{r} + O(\rho)$

With the size of dominating set, we can express T_{cost} as:
$E[T_{\text{cost}}] = \frac{L^2}{\rho^2} + (|R(\rho)| - 1)(4\frac{L}{r} + O(\rho))$

The cost of transmission per broadcast then:

$$E[E_{\text{cost}}] = \frac{L^2}{\rho^2(|R(\rho)| - 1)} + \frac{4L}{r} + O(\rho)$$

Let us consider the ratio of E_{cost} with the transmission-level bound E_{bound} from section 2.1.2, $E_{\text{bound}} = \frac{L^2}{\pi r^2}$

Considering the limit of all involved quantities, we get $\frac{1}{\text{Eff}_{\text{bound}}} = \frac{E_{\text{cost}}}{E_{\text{bound}}} \to 1 + \frac{4\pi r}{L}$ when $\mu \to \infty$ conditioned to the event $\mathcal{E}$ whose probability $Pr[\mathcal{E}] \to 0$

5.7 Proof of the Main Discrete Geometry Property on Neighborhood

Theorem 5. *If X and Y are a partition of the integer lattice $\mathbb{Z}^n$, and X is finite, then the following two properties (12) and (13) are verified:*

$$|\text{border}(X)| \geq M_R, \text{ or else : border}(X) = X \tag{12}$$

$$|\text{border}(Y)| \geq M_R \tag{13}$$

where R is the set of neighbors of the origin node, and $M_R = |R| - 1$

5.7.1 Preliminaries

The proof is based on the use of the Minkowski addition, and a specific property of discrete geometry (15) below. The Minkowski addition is a classical way to express the neighborhood of one area

Given two sets A and B of $\mathbb{R}^n$, the Minkowski sum of the two sets $A \oplus B$ is defined as the set of all vector sums generated by all pairs of points in A and B, respectively:

$$A \oplus B \triangleq \{a + b \mid a \in A, b \in B\} \tag{14}$$

Then the closed set of neighbors $\mathcal{N}[t]$ of one node t, can be redefined in terms of Minkowski sum: $\mathcal{N}[t] = \{t\} \oplus R$.
This extends to the neighborhood of subsets: $\mathcal{N}[A] = A \oplus R$

For Minkowski sums on the lattice $\mathcal{L}$, there exist variants of the *Brunn-Minkowski inequality*, including the following one [18]: For two non-empty subsets A, B of the integer lattice $\mathbb{Z}^n$,

$$|A \oplus B| \geq |A| + |B| - 1 \tag{15}$$

5.7.2 Proof for Th. 5

Proof. We start by proving (12) of Th. 5

Consider the set inner(X). If it is empty, by definition of inner(X), all nodes of X must be neighbors of nodes of Y: $X = \text{border}(X)$. This implies (12) of Th. 5.

Otherwise, inner$(X) \neq \emptyset$: Again by definition we have:

$$\begin{aligned}
\mathcal{N}[\text{inner}(X)] &\subset X \\
\mathcal{N}[\text{inner}(X)] \setminus \text{inner}(X) &\subset X \setminus \text{inner}(X) \\
\mathcal{N}[\text{inner}(X)] \setminus \text{inner}(X) &\subset \text{border}(X) \quad \text{(by def. of border)} \\
\text{Therefore}: \ |\text{border}(X)| &\geq |\mathcal{N}[\text{inner}(X)] \setminus \text{inner}(X)|
\end{aligned}$$

For the second part of this inequality, we have:

$$\begin{aligned}
|\mathcal{N}[\text{inner}(X)] \setminus \text{inner}(X)| &\geq |\mathcal{N}[\text{inner}(X)]| - |\text{inner}(X)| \\
&\overset{(a)}{\geq} |\text{inner}(X) \oplus R| - |\text{inner}(X)| \\
&\overset{(b)}{\geq} |\text{inner}(X)| + |R| - 1 - |\text{inner}(X)|
\end{aligned}$$

(a) is by rewriting neighborhood with a Minkowski sum, (b) is obtained by using (15).

Hence $|\text{border}(X)| \geq |R| - 1$ which implies (12) of Th. 5, since $M_R = |R| - 1$ by definition.

• For (13) of Th. 5: By definition, border(Y) includes all nodes $y \in Y$ that are neighbors of nodes of X. Hence:
$\mathcal{N}(X) \subset \text{border}(Y)$
$\mathcal{N}[X] \setminus X \subset \text{border}(Y)$ by def. of $\mathcal{N}[X]$
Hence: $|\text{border}(Y)| \geq |\mathcal{N}[X] \setminus X|$
The second part of the equation can be written:

$$\begin{aligned}
|\mathcal{N}[X] \setminus X| &\geq |\mathcal{N}[X]| - |X| \\
&\overset{(a)}{\geq} |X \oplus R| - |X| \\
&\overset{(b)}{\geq} |X| + |R| - 1 - |X|
\end{aligned}$$

(a) is by rewriting neighborhood with a Minkowski sum, and (b) is with inequality (15). As a result $|\text{border}(Y)| \geq |R| - 1$, which is Eq.(13) of Th. 5. □

6 Conclusion

We have presented methods of rate selection for network coding in multi-hop wireless networks. The logic behind these rate selections was described. The methods are based on the use of dominating sets: to compensate for the effect of the borders (nonexistent neighbors); to compensate for occasional lack of neighbors (holes); and to adjust for non-uniform density of the network.

Proofs were given for their performance, and rely on the proof of maximum broadcast rate of the source. We proved that RAUDS will achieve optimal energy-efficiency locally and that MARAUDS would asymptotically achieve optimal energy-efficiency when the area and density of the network would grow indefinitely. This was used to derive a upper and lower bound of the gain of network coding in general compared to routing under these hypothesis: between 1.642 and 1.684 (in Eq. (7)).

For dense networks in a fixed square region of the plane, an estimate was given for the condition for expecting the rate selection MARAUDS to outperform any method using routing: it is the case when the radio range smaller that one $\frac{1}{20^{th}}$ of the edge length of the square. Future research work will explore behavior on less dense graphs, the practical performance of the methods, and choices of parameters.

References

[1] R. Ahlswede, N. Cai, S.-Y. R. Li and R. W. Yeung, *"Network Information Flow"*, IEEE Trans. on Information Theory, vol. 46, no.4, Jul. 2000

[2] S. Guha and S. Khuller, *"Approximation algorithms for connected dominating sets"*, European Symposium on Algorithms, 1996.

[3] B. Das and V. Bharghavan *"Routing in Ad-Hoc Networks Using Minimum Connected Dominating Sets"*, ICC 1997.

[4] F. Dai and J. Wu, *"An Extended Localized Algorithm for Connected Dominating Set Formation in Ad Hoc Wireless Networks"*, IEEE Transactions on Parallel and Distributed Systems, v.15 n.10, p.908-920, Oct. 2004

[5] C. Fragouli, J. Widmer, and J.-Y. L. Boudec, *"A Network Coding Approach to Energy Efficient Broadcasting"*, INFOCOM 2006, Apr. 2006

[6] J.-S. Park, M. Gerla, D.S. Lun, Y. Yunjung Yi and M. Médard *"Codecast: a network-coding-based ad hoc multicast protocol"* IEEE Wireless Communications, IEEE Volume 13, Issue 5,Oct. 2006

[7] Y. Wu, P. A. Chou, and S.-Y. Kung, *"Minimum-Energy Multicast in Mobile Ad Hoc Networks using Network Coding"*, IEEE Trans. Commun., vol. 53, no. 11, pp. 1906-1918, Nov. 2005

[8] D. S. Lun, N. Ratnakar, M. Médard, R. Koetter, D. R. Karger, T. Ho, E. Ahmed, and F. Zhao, *"Minimum-Cost Multicast over Coded Packet Networks"*, IEEE/ACM Trans. Netw., vol. 52, no. 6, pp 2608-2623, Jun. 2006

[9] J. Y. Yu and P. H. J. Chong, *"A survey of clustering schemes for mobile ad hoc networks"*, IEEE Communications Surveys & Tutorials, First Qtr. 2005

[10] B. Clark, C. Colbourn, and D. Johnson, *"Unit disk graphs"*, Discrete Mathematics, Vol. 86, Issues 1-3, Dec. 1990

[11] D. S. Lun, M. Médard, R. Koetter, and M. Effros, *"On coding for reliable communication over packet networks"*, Tech. Report #2741, MIT LIDS, Jan. 2007

[12] T. Ho, R. Koetter, M. Médard, D. Karger and M. Effros, *"The Benefits of Coding over Routing in a Randomized Setting"*, International Symposium on Information Theory (ISIT 2003), Jun. 2003

[13] D. S. Lun, M. Médard, R. Koetter, and M. Effros, *"Further Results on Coding for Reliable Communication over Packet Networks"* International Symposium on Information Theory (ISIT 2005), Sept. 2005

[14] N. Karmarkar, *A new polynomial-time algorithm for linear programming*, Combinatorica, 4:373-395, 1984.

[15] A. Ramamoorthy, J. Shi, and R. D. Wesel, *"On the Capacity of Network Coding for Random Networks"*, IEEE Trans. on Information Theory, Vol. 51 No. 8, pp. 2878-2885, Aug. 2005

[16] S. A. Aly, V. Kapoor, J. Meng, and A. Klappenecker, *"Bounds on the Network Coding Capacity for Wireless Random Networks"*, 3rd Workshop on Network Coding, Theory, and Applications (Netcoding07), Jan. 2007

[17] M. Jafarisiavoshani, C. Fragouli, and S. Diggavi *"Subspace Properties of Randomized Network Coding"* Information Theory for Wireless Networks, ITW 2007, Jul. 2007

[18] R. J. Gardner, P. Gronchi, *"A Brunn-Minkowski inequality for the integer lattice"*, Trans. Amer. Math. Soc., 353 (2001), 3995-4042

[19] K. Voss, *Discrete Images, Objects, and Functions in* Z^n, Springer-Verlag 1993.

[20] C. Adjih, S. Y. Cho and P. Jacquet, *"Near Optimal Broadcast with Network Coding in Large Sensor Networks"*, Proceedings of the first workshop on Information Theory for Sensor Networks, june 2007.

[21] C. Adjih, S. Y. Cho and P. Jacquet, *"Near Optimal Broadcast with Network Coding in Large Homogeneous Networks"*, INRIA Research Report RR-6188, may 2007.

[22] C. Adjih, S. Y. Cho, manuscript notes of this technical report, `WNC-CDS-manuscript-notes.pdf`, available from at `hal.inria.fr` in the source distribution of this report.

Centre de recherche INRIA Paris – Rocquencourt
Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Centre de recherche INRIA Bordeaux – Sud Ouest : Domaine Universitaire - 351, cours de la Libération - 33405 Talence Cedex
Centre de recherche INRIA Grenoble – Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier
Centre de recherche INRIA Lille – Nord Europe : Parc Scientifique de la Haute Borne - 40, avenue Halley - 59650 Villeneuve d'Ascq
Centre de recherche INRIA Nancy – Grand Est : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex
Centre de recherche INRIA Rennes – Bretagne Atlantique : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex
Centre de recherche INRIA Saclay – Île-de-France : Parc Orsay Université - ZAC des Vignes : 4, rue Jacques Monod - 91893 Orsay Cedex
Centre de recherche INRIA Sophia Antipolis – Méditerranée : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399