Network Coding in Planar Networks

Tang Xiahou†, Zongpeng Li‡‡, Chuan Wu§
† Department of Computer Science, University of Calgary, Canada
‡ Institute of Network Coding, Chinese University of Hong Kong
§ Department of Computer Science, The University of Hong Kong
{txiahou, zongpeng}@ucalgary.ca, cwu@cs.hku.hk

Abstract—A basic problem in network coding is to choose a field to perform the encoding and decoding operations in. The main message we wish to deliver in this work is that while unboundedly large fields are required for linear algebraic codes in arbitrary networks, very small finite fields may be sufficient for networks in practice. We use planar networks as a representative example of a practical network, which usually has a linear number of links, and exhibits a topology that is planar or close to planar. Focusing on a basic scenario of multicasting two source information flows, we prove that coding over $GF(3)$ is sufficient for all planar networks. The proof applies existing techniques in the literature, including multicast flow decomposition, graph coloring, as well as new techniques that made these tools readily applicable to all undirected planar multicast networks. We further show that $GF(3)$ is also necessary, by constructing two new planar multicast networks, one of which is a bipartite, that require coding over $GF(3)$ for achieving the maximum throughput. For the special case of outerplanar networks, we prove that network coding is always equivalent to routing. Our results also invite a renewed look at the choice of deterministic vs. randomized network coding in practical networks.

I. INTRODUCTION

Among basic questions in network coding, such as where and how to encode [1]–[3], how large is the gap between network coding throughput and routing throughput [4]–[6] and when a large fraction of network codes need to encode [7], the choice of a finite field to perform the encoding and decoding operations in is also fundamental. It is known that for a single multicast session in a directed network, the minimum of the min-cut values between the source and each receiver can be achieved for all receivers concurrently, if network coding over a sufficiently large field is allowed [1]. But how large is large enough? The required field size grows as the network size grows, and no constant size is always sufficient. For example, the combination network $C_{n,k}$ [6] requires a field size at least $n-1$. The best known result so far, on sufficiently large fields, is based on the observation that encoding is necessary only when flows to different receivers merge. The number of flows merging at a single link is upper-bounded by the number of receivers, and it is always sufficient to have the field size $q$ to be the number of multicast receivers [3], [8].

A problem of using a very large base field is the high computational overhead in encoding and decoding operations. That is on top of the complexity of code assignment — deterministic code assignment algorithms designed for arbitrary networks also incur a relatively high overhead, with polynomial but not linear time complexity [3].

Very large fields and generic code assignment algorithms may indeed be an overkill for networks in practice. Real-world computer networks deployed today exhibit two important features. First, the average node degree is constant and does not grow as the network size grows. Consequently, the network contains a linear number of links. Second, the underlying topology of the network is often planar or close to planar. A planar graph is one that can be drawn on a piece of paper without two edges crossing each other. As shown in Fig. 1, metropolitan and backbone networks are deployed along the surface of the globe, showing a natural plane embedding. Euler’s formula further reveals that, a planar network of $n$ nodes has at most $3n-6$ links, which is linear [9].

Another probably more important motivation for studying planar networks, from a networking perspective, is that planar networks are well understood, have good properties, and allow extremely efficient protocols to be designed and executed. A series of recent work in the networking literature propose to construct a planar network backbone from a non-planar network, such as a densely deployed wireless sensor network, for running network protocols such as broadcast and multicast, in a more efficient and effective manner (e.g., [10], [11]).

We choose planar networks as a representative example of networks from practice, and focus on a basic communication scenario of multicasting two source information flows to a set of receivers. A simple example of such a network is the combination network $C_{3,2}$, shown in Fig. 2. We prove that coding over $GF(3)$ is sufficient in all undirected multicast networks. Our study has been inspired by the following two
closely related work. Fragouli et al. proposed the technique of decomposing a multicast flow enabled by network coding into subtree components, and suggested that the code assignment problem can be transformed into a version of subtree coloring problem [12]. El Rouayheb et al., in their study of minimal directed multicast networks, showed that coding over $GF(3)$ is sufficient for 2-minimal directed acyclic multicast networks [13]. We combine the multicast flow decomposition technique [12] and the idea of four-coloring a planar graph [13], together with new techniques that make these tools readily applicable to all planar networks. Our proof also shows that code assignment over $GF(3)$ can be performed in linear time in a planar network — a consequence of the fact that a planar graph can be four-colored in linear time [14].

Our results show that in planar networks, deterministic code assignment requires a small field $GF(3)$, and has linear time complexity. In comparison, randomized network coding [15] also has linear aggregated complexity (constant time per node). However, our simulation results reveal that a relatively large field is still required for randomized network coding to have a high success rate in a planar network, and $GF(3)$ is far from satisfactory. While deterministic network coding compares unfavorably to randomized network coding in general networks due to its higher complexity, the picture charges sharply in planar networks. Randomized network coding appears incapable of exploiting the planar structure of the network.

To our knowledge, all existing multicast network patterns in the literature, if planar, require coding over $GF(2)$ or no coding at all. For example, Fig. 3 [7] shows a network pattern that is planar, and generalizes the butterfly network from throughput $h = 2$ to $h = 4$. Although it requires coding at many nodes [7], it requires only the binary field $GF(2)$ for performing coding operations.

Nonetheless, we show that the bound 3 is tight, i.e., a field of size 3 is necessary for network coding in planar networks in general. We construct two new multicast networks, each requiring coding over $GF(3)$. The first is obtained by connecting two juxtaposed copies of the butterfly network, which has an isomorphic topology to the combination network $C_{3,2}$. The second is constructed by conducting surgeries on the combination network $C_{4,2}$, which requires $GF(3)$ but is not planar, into a planar network. The bipartite property of $C_{4,2}$ is kept intact in the second network.

A special case of planar networks is an outerplanar network, which has a face that is adjacent to all nodes. In graphical illustrations showing a plane embedding, this face is usually chosen to be the outer infinite face (a planar graph can be embedded with any face as the infinite face). The classic butterfly network [1] is a well known example that requires network coding for achieving the maximum multicast throughput. The butterfly network is planar, but not outerplanar. As shown in Fig. 4, if we contract the bottleneck link, the resulting 6-node 8-link network becomes outerplanar, and does not require network coding anymore. We show that this is not a coincidence, by proving that for all outerplanar networks, network coding is equivalent to routing. The proof techniques are similar to those applied for showing $GF(3)$ is sufficient in all general planar networks. A general planar graph requires four colors for coloring, and we indeed used four ‘colors’, representing four pair-wise linearly independent combinations (over $GF(3)$) of two source flows. An outerplanar graph requires three colors for proper coloring, but we have only two: the two source flows only. We show how this barrier can be cleared.

While results in this paper are confined to multicasting two flows, we suspect that the same hold in general settings with larger throughputs. In other words, $GF(3)$ is sufficient for network coding in planar networks in general, and network coding is equivalent to routing in outerplanar networks in general. The proofs to the general case are left as future work.

II. NETWORK MODEL

We consider an undirected multicast network $G = (V, E)$. Links in $E$ are capacitated, and capacities are stored in an integer vector $c$. A source $S$ wishes to multicast $h$ source information flows to $k$ receivers $T_1, \ldots, T_k$. In this paper, we focus on the basic case of $h = 2$. The multicast source and receivers are jointly referred to as the terminal nodes.

Fig. 2. The combination network $C_{3,2}$ is a planar network. Maximum throughput 2 can be achieved by coding over $GF(2)$. Maximum throughput with routing (tree packing) is 1.8 [6]. Coding advantage is $2/1.8 = 10/9$.

Fig. 3. A generalized version of the butterfly network, planarity maintained, with throughput $h = 4$. Required field is $GF(2)$. Coding advantage is $16/15 = 1.067$. Around one third of nodes need to perform encoding operations, as the pattern grows [7].

Fig. 4. Left: the butterfly network, shown in undirected form, is a planar multicast network. Coding over $GF(2)$ is necessary. Coding advantage is $2/1.875 = 1.067$. Right: an outerplanar multicast network obtained by contracting the bottleneck link in the butterfly network into a node. Coding is not necessary.
Other nodes in $G$ are relay nodes. We assume the network has sufficient capacity to support throughput $h = 2$, with a static linear algebraic code. In other words, the network and the multicast flow may contain cycles, but we assume the concept of time does not have to be involved, and a linear convolutional code [16] is unnecessary, for ease of analysis. We leave as future work the more general study where the multicast throughput can only be achieved using a linear convolutional code.

As shown in Fig. 5, a link $e$ of capacity $c(e)$ can be replaced with $c(e)$ unit-capacity links. Since we consider the case of $h = 2$, it is sufficient to consider $c(e) \in \{1, 2\}$, and there are at most two parallel links between a pair of neighbor nodes.

![Fig. 5. Converting a capacitated link to parallel unit-capacity links.](image)

The coding advantage of a multicast network $G$ is the ratio of the maximum throughput achievable with network coding over that with routing (tree packing). For planar networks shown in figures throughout this paper, we list the coding advantage computed, by solving a maximum multicast throughput LP and a tree packing LP [17], and make a conjecture in the conclusion section accordingly.

III. The Sufficiency of $GF(3)$ in General Planar Networks

In this section, we prove the sufficiency of coding over $GF(3)$ in a general planar network.

**Theorem 1.** Coding over $GF(3)$ is sufficient for multicasting two information flows in an undirected planar network.

**Proof:** We present a constructive proof to the theorem, by performing code assignment using $GF(3)$ only, in the following three steps.

1. **Multicast flow decomposition.** In this step, we first compute a multicast flow $f$ (a vector with a non-negative flow rate on each link) achieving throughput two [17]. We perform code assignment over this flow $f$. To start, we apply the subtree decomposition technique proposed by Fragouli et al. [12], and decompose $f$ into a set of subtree components with the following properties: each component is a directed tree (subtree) with a single root node; the root and leaves of each subtree have incoming degree 2; no two subtrees intersect with each other. In our case, where the original network $G$ is planar, each subtree component further induces a face around it, and a plane embedding of $G$ is thus decomposed into subtree faces. Fig. 7 shows a multicast flow decomposition, where the network $G$ is decomposed into four faces, each shown in a distinct color.

![Fig. 7. Directions of faces with respect to nodes. The subtree faces can always be colored using four different colors, such that no two adjacent faces share a common color.](image)

To help understanding the concept of subtree decomposition, one can imagine a valid code assignment over $f$ using a sufficiently large field. Then the multicast flow can be decomposed into maximally connected trees, each containing links transmitting a common information flow. In cases where two common-flow trees intersect, they need to be further decomposed.

2. **Introducing new face boundaries.** A boundary node in the subtree decomposition is a node that lies at the boundary of more than one faces. It must be a root or a leaf of a subtree. With respect to a given boundary node $u$, we say an adjacent face is an in-face if $u$ is a leaf in that face, and an out-face if $u$ is a root of that face.

We perform the following surgery on the decomposed subtree faces. If a boundary node of a subtree has degree 2, and has two flows arriving from two opposite faces, then we expand this node into a node pair, connected by a new link, as shown in Fig. 6. Such a link expansion is preparing for the four-coloring of the subtree faces, as will be evident in the next step.

![Fig. 6. Expanding a node into two nodes connected by a link, while preserving the planarity of the network. The two opposite in-faces have a new common boundary due to the expansion, and will be assigned different colors during the four-coloring process.](image)

3. **Code assignment via face coloring.** By the Four Color Theorem in graph theory [9], every planar graph can be colored using four colors, such that no two adjacent faces share a common color. Two faces are adjacent if they share a common link, and are not adjacent if they share only a common node. Assume the two original source flows are denoted as $x$ and $y$, each at unit rate. We color the four faces resulting from our multicast flow decomposition in step 2, and then processed in step 3, using the following four colors: $x$, $y$, $x + y$ and $x + 2y$. We claim that such a coloring leads to a proper code assignment to the multicast flow $f$.
root $u$ of each subtree face has two distinct incoming flows. This is true since $u$ has two in-faces due to the property of the subtree decomposition. Furthermore, the expansion operation in Step 3 guarantees that these two in-flows always share a common boundary, and hence will be assigned different colors/flows. Once receiving two distinct flows, the root $u$ is able to linearly combine them for generating the flow assigned to its out-face(s), if any. To conclude, we obtained a valid code assignment over $GF(3)$, for achieving multicast throughput $h = 2$ in the planar network.

As previously mentioned, our proof has been inspired by both the work of Fragouli et al. [12] on multicast flow decomposition, and by the work of El Rouayheb et al. [13] on using four linearly independent flows to four-color a planar network. The result we obtain is similar to that of El Rouayheb et al., though our result is slightly general in that it is regarding undirected networks admitting a static code. Our proof is also somewhat different from that of El Rouayheb et al., and we hope the new proof structure is easier to be adapted to more general settings such as when $h > 2$. In Sec. V, we apply our proof technique here to further show that network coding is not necessary in outerplanar networks.

Although the current result is for $h = 2$ only, we suspect that $GF(3)$ is sufficient for all undirected planar multicast networks, with $h \geq 2$. Robertson et al. [14] show that the four coloring of a planar graph can be done in time linear to the number of nodes in the graph. In comparison, randomized network coding also takes linear aggregated time across the network (constant time per node). While a deterministic code assignment algorithm can exploit the underlying planarity of the network and operate over a small field $GF(3)$, it is unclear how a randomized code assignment algorithm performs, in terms of success rate in delivering independent flows to all multicast receivers, in planar networks over small fields. We have conducted simulation studies, where we use plantri [18] to generate random planar networks with 10 nodes, and set $h = 2$. A multicast flow of rate 2 is computed, and randomized linear combinations are performed at nodes with in-degree 2 who have downstream neighbors.

The table below shows the relation between success probability and field size. For simplicity, we choose prime fields for testing. Each success rate is the average over three testings in different networks. As we can see, the success rate is rather poor for small fields with size 2 or 3, and remains unsatisfactory up to 23. Only when the field size grows to 131 and 311, does the success rate improve to 0.979 and 0.991, respectively. These results appear similar to the performance of randomized network coding in general networks that is not necessarily planar.

<table>
<thead>
<tr>
<th>field size</th>
<th>2</th>
<th>3</th>
<th>5</th>
<th>7</th>
<th>11</th>
<th>23</th>
<th>131</th>
<th>311</th>
</tr>
</thead>
<tbody>
<tr>
<td>success rate</td>
<td>0.296</td>
<td>0.423</td>
<td>0.582</td>
<td>0.670</td>
<td>0.770</td>
<td>0.881</td>
<td>0.979</td>
<td>0.991</td>
</tr>
</tbody>
</table>

IV. THE NECESSITY OF $GF(3)$ IN GENERAL PLANAR NETWORKS

While Theorem 1 in Sec. III shows that $GF(3)$ is sufficient in planar networks, we next show that it is also necessary in general. We construct two specific undirected multicast networks, where the maximum multicast throughput can be achieved by coding over $GF(3)$ but not over $GF(2)$.

**Theorem 2.** Coding over $GF(3)$ is necessary for multicasting 2 source flows in general planar networks.

*Proof:* Fig. 8 depicts a multicast network (the same network as used in Fig. 7), with one source $S$ and five receivers $\{T_1, T_2, T_3, T_4, T_5\}$. Each link has a unit capacity. To achieve throughput 2, there is only one possible orientation and one feasible multicast flow (shown). Observing that it requires at least 4 linear independent flows to satisfy the demands of receivers $T_i, i = 1, 2, 3, 4, 5$, so a minimum field size of 3 is necessary in this network.

![Fig. 8. A planar multicast network constructed from two copies of $C_{3,2}$. Optimal throughput 2 can be achieved by coding over $GF(3)$, as shown, but cannot be achieved by coding over $GF(2)$. Coding advantage is $2/1.882 = 1.063$.](image)

**Theorem 3.** Coding over $GF(3)$ is necessary for multicasting 2 source flows in bipartite planar networks.

*Proof:* Fig. 9 shows a multicast network with one source, four relay nodes and six receivers, similar to those in the combination network $C_{4,2}$. Further resembling $C_{4,2}$ is the fact that $GF(3)$ is required for achieving a multicast throughput of 2. The underlying reason for the necessity is also exactly the same as in $C_{4,2}$: each of the $\binom{4}{2} = 6$ receivers is connected to a distinct set of the 4 relays, via two unit capacity links. Therefore each relay node needs to receive and forward a different flow, such that the four flows together are pair-wise linearly independent, requiring a field size of at least 3. While $C_{4,2}$ is not planar, the network in Fig. 9 is. Both are also bipartite networks.

While the maximum throughput with network coding is 2 in both networks, tree packing can achieve a higher throughput in the network in Fig. 9. The network $C_{4,2}$ has a coding advantage of 9/8, which is the largest among all combination networks (tie with $C_{4,3}$) [6], and the largest among all small to moderate size multicast networks known. The network in Fig. 9 has a coding advantage of 10/9, which is slightly smaller, and is the largest value known in undirected planar networks (tie with $C_{3,2}$).
From Theorem 2, we can derive the following corollary:

**Corollary 1.** Linear algebraic code over \( GF(2) \) with half-integer routing is sufficient for multicasting two source flows in planar networks.

**Proof:** The proof is basically a combination of Theorem 2 with the fact that linear block codes can 'mimic' linear algebraic codes in the multicast setting [16]. More specifically, given any valid code assignment over \( GF(3) \) with integral flows, we can perform a universal transformation and derive a code assignment over \( GF(2) \) with half-integral flows. Divide the two original unit rate flows \( x \) and \( y \) into two 0.5-rate flows respectively, \( \{x_1, x_2\} \) and \( \{y_1, y_2\} \). In the original code assignment over \( GF(3) \), replace \( x \) with \( \{x_1, x_2\} \), \( y \) with \( \{y_1, y_2\} \), \( x + y \) with \( \{x_1 + y_1, x_2 + y_2\} \), and \( x + 2y \) with \( \{x_1 + x_2 + y_2, x_2 + y_1\} \). \( \square \)

V. THE EQUIVALENCE BETWEEN NETWORK CODING AND ROUTING IN OUTERPLANAR NETWORKS

Recall that when we transform the butterfly network into an outerplanar multicast network without the bottleneck link (Fig. 4), network coding is not necessary any more. We prove that this is not a coincidence. Note that star networks, trees and cycles are special cases of outerplanar networks. The links in an outerplanar network can be divided into two categories: outer links on the infinite face, and inner links (chords).

**Theorem 4.** In an outerplanar multicast network, with \( h = 2 \), network coding is equivalent to routing.

**Proof:** We present a constructive proof to the theorem, by designing a routing solution (code assignment using the two original flows \( x \) and \( y \) only) in the following four steps.

1. **Multicast flow decomposition.** Similar to the proof of Theorem 1, we first compute a multicast flow \( f \) achieving the throughput 2, and decompose \( f \) into a set of subtree faces.

2. **Forming two regions.** Given the subtree face graph from Step 1, we traverse all boundary nodes (root and leaves of each subtree). If a boundary node \( u \) has two in-faces only (corresponding to a receiver in the multicast flow without outgoing flows), we merge its two adjacent faces \( F_\alpha \) and \( F_\beta \) into a new face \( F_\alpha \cup F_\beta \), as shown in Fig. 10. Later on we will color the subtree faces. A proper coloring to the merged faces can be easily converted to a coloring of the original subtree faces: let \( F_\beta \) inherit the color of \( F_\alpha \cup F_\beta \), then pick the complementary color for face \( F_\alpha \), who has only one neighbor face.

After the above face merging, we traverse all boundary nodes for a second round. For each boundary node \( u \), we label the two of \( u \)'s adjacent faces that are neighboring the infinite face as **region 1 faces**, as shown in Fig. 11. All links not included in region 1 (must be chords) are in **region 2**. We next color the two regions separately, using \( x \) and \( y \).

3. **Coloring region 1.** We color the faces in region 1 along the infinite face boundary using two colors \( x \) and \( y \). A potential conflict arises due to the availability of two colors only, if the number of region 1 faces is odd, and each of them share a expanded boundary with each of its two neighbor faces. We show that this is impossible.

As shown in Fig. 12, if \( F_R \) is an in-face with respect to node \( v_1 \), then it must be an out-face for \( v_2 \). The reason is that \( F_R \) has only two boundary nodes, for one of which it must be an out-face. Consequently, terminal node \( v_2 \) would not have been expanded, and its two adjacent region 1 faces have no conflict.

4. **Coloring region 2.** We next color chord links in region 2. As shown in Fig.13, a chord in region 2 cannot enter an expanded boundary node — otherwise that boundary node would have three incoming flows, contradicting \( h = 2 \). When the chord enters a non-expanded boundary node \( u \), \( u \) has one in-face only in region 1, and we can pick the complementary color to that in-face for coloring the chord link.
To conclude, we have obtained a valid code assignment for the multicast flow $f$ using two original flows $x$ and $y$ only, for the outerplanar multicast network.

An interesting middle ground between general planar multicast networks and outerplanar multicast networks are terminal-co-face multicast networks, which are planar, and have all terminal nodes lying on a common face. For examples, refer to the butterfly network in Fig. 4 and the generalized butterfly network in Fig. 3. In both examples, the binary field $GF(2)$ is sufficient. We conjecture that this is not a coincidence.

**Conjecture 1.** Coding over $GF(2)$ is sufficient for terminal-co-face multicast networks.

**VI. CONCLUSION AND OPEN PROBLEMS**

In this work, we studied network coding in planar networks, focusing on the basic case of multicasting two source flows. Our results reveal that $GF(3)$ is sufficient and necessary for general planar networks, and that network coding is not necessary in outerplanar networks. We conjecture that $GF(2)$ is sufficient for the middle ground of terminal-co-face networks. A number of problems are open for future research.

First, we conjecture that the results proven in this work hold not only for the case of $h = 2$, but also for any $h \geq 2$ in general. Second, the best known solution for computing the maximum multicast throughput in undirected networks involves solving a linear program [17]. It is interesting to study whether a strongly polynomial time algorithm can be designed, without requiring solving LPs, for undirected planar networks. Third, all observed coding advantage values in undirected planar multicast networks are at most 10/9. Is 10/9 indeed a tight upper-bound?

It is known that a graph is planar if and only if it does not contain as a minor $K_5$ or $K_{3,3}$, and is outerplanar if and only if it does not contain as a minor $K_4$ or $K_{3,2}$ [9]. It appears interesting to study network coding from the graph minor perspective — what is the fundamental structure in the network topology that enforces network coding, and coding over a large field? We make the following conjectures along this direction: (1) a directed multicast network requires network coding only if it contains a $K_4$ minor, (2) an undirected multicast network requires network coding only if it contains a $C_{3,2}$ minor, (3) an undirected multicast network requires $GF(2^2)$ only if it contains a $K_5$ minor, and more generally, (4) there exists a function $f(q)$, such that if an undirected multicast network $G$ requires coding over $GF(q)$, then $G$ must contain a minor $K_{f(q)}$, and $f(2) = f(3) = 4$, $f(4) = 5$. All results we proved and conjectured previously in this paper can be viewed as corollaries of these propositions from the graph minor perspective.

**REFERENCES**