On Optimality of Routing for Multi-source Multicast Communication Scenarios with Node Uplink Constraints

Sudipta Sengupta* Minghua Chen† Philip A. Chou* Jin Li*
*Microsoft Research, Redmond, WA, USA. Email: {sudipta,pachou,jinl}@microsoft.com
†The Chinese University of Hong Kong, Shatin, NT, Hong Kong. Email: minghua@ie.cuhk.edu.hk

***Abstract*—We consider multi-source multicast communication scenarios in which each node has an aggregate outbound traffic capacity and can directly communicate with any other node. This is motivated by peer-to-peer (P2P) information dissemination applications on the Internet in which the uplink capacity of nodes is usually the bottleneck, being several times smaller than the downlink capacity. We also allow the communication in a group to be helped by non-receiver nodes (with respect to that group) as relays. Extending an earlier result for the single source case, we show that when coding is not allowed across sources, routing is optimal. Also, as a rather surprising discovery, we show that when all groups have pairwise identical or disjoint receivers, routing is optimal even when coding across sources is allowed. Moreover, routing along a linear number of trees per source is sufficient to achieve this. The latter scenario is common in multi-party conferencing systems, hence our results have interesting practical applications in the design of infrastructure-less P2P multiparty conferencing systems.**

I. Introduction

The achievable rate region of single source multicast scenario is characterized as the minimum of the min-cuts between the source node s and all nodes in its receiver set R [1], i.e., $\min_{t \in R}$ min-cut (s,t). If network coding is allowed, then the single source multicast rate region can be achieved for arbitrary topology by solving the routing and coding problems separately, each being of polynomial complexity [4].

The achievable rate region for multi-source multicast scenarios was recently implicitly characterized in [7], but currently no scheme is known to achieve it. It is believed that information from different multicast groups should be coded in a nonlinear fashion in order to achieve the rate region (inter-session coding). However, doing such mixing and coding is complex and largely an open problem.

Regardless of its power, network coding is not quite practical in many applications today. It cannot be used in the Internet routing layer because it requires changes in all routers (for encoding) and end-hosts (for decoding). If deployed in the overlay layer, it will introduce new complexity in end-host software (for encoding and decoding) and additional delays in packet delivery.

A practical way to explore the achievable rate region is by routing. Each source s packs directed (Steiner) trees rooted at s and reaching all its receivers R_s. For the *general case of arbitrary topologies*, this approach of routing brings up the following difficulties:

1) For a given source, the maximum rate achieved by routing can be a factor of up to $\log n$ lower than that achieved by network coding [4], where n is the number of nodes.
2) To achieve the maximum rate for routing, the problem of packing directed Steiner trees is $\mathcal{NP}$-hard [5]. Moreover, the number of Steiner trees used in an optimal solution may be exponential.
3) For the special case when source and receivers comprise all nodes in the network), Edmonds' theorem [2] states that the min-cut bound for source s and receivers R can be achieved by packing directed spanning trees (arborescences) [3]. Moreover, an optimal packing can be determined in polynomial time and uses at most m distinct arborescences [3], where m is the number of edges (which specific arborescences are used in an optimal solution depends on the link capacities). However, the problem we consider has multiple sources as well as relay nodes, hence Edmonds' theorem does not directly help to establish any optimality results for routing.

As such, routing can not achieve the optimal rate region in general topology and its cost could be prohibitively large.

The problem that we consider in this paper is motivated by Peer-to-Peer (P2P) applications that have witnessed unprecedented growth on the Internet in recent years and are increasingly being used for real-time applications like video conferencing and live streaming. The fact that peer node uplinks are the only bottlenecks (in practice) in the network for such applications allows us to tackle all of the above difficulties with routing as well as establish its optimality in a surprisingly elegant manner. In the remainder of this section, we describe our assumptions for what we call a *P2P topology*, then move on to some notation, summarize our contributions, and describe P2P application scenarios for our results.

A. P2P Topology

The term "P2P topology", as used in this paper, is an *overlay network topology*, consisting of end-hosts (peers) as nodes and connections between them as edges (possibly realized through routing on the public Internet as the underlay) with the following properties:

- Each node has an uplink capacity, and these uplinks are the only rate limiting bottlenecks in the whole network, and
- Every node can (possibly) communicate directly with every other node, subject to peer resource and policy constraints.

In the overwhelming majority of residential broadband connections, bottlenecks typically are at the edge of the access networks rather than in the middle of the Internet.

Furthermore, it is common to have the uplink capacity of a peer to be several times smaller than the downlink capacity, thus justifying the practicality of our assumption on P2P topology. Formally, if a peer i has uplink capacity $c_{out}(i)$, downlink capacity $c_{in}(i)$, and is a source of data at rate R_i, and a sink of data at rate R'_i (i.e., it is not uploading this data to any other peer), then its downlink is not a bottleneck if $c_{in}(i) \geq R'_i + (c_{out}(i) - R_i)$.

B. Notation

Let $G = (V, E)$ be the underlying communication (directed) graph. Let $S \subseteq V$ be a designated set of senders, and for each $s \in S$, let $R_s \subseteq V$ be a designated set of receivers (including s itself). An information source originates from each sender $s \in S$, and is to be received by each $r \in R_s$. Let the remaining nodes $H_s = V - R_s$ be a set of helpers that can act as relays for distributing the information from source s. Each helper $h \in H_s$ may use its upload bandwidth to help distribute information from senders to receivers. *The graph G is a complete (full-mesh) graph except for the scenario in Section III where we do not allow an edge between two receivers if they receive content from different sets of sources.*

Let $In(v)$ denote the set of edges entering node v and let $Out(v)$ denote the set of edges leaving node v. Let $c_{out}(v)$ denote the upload capacity (i.e., output bandwidth) of each node v in V.

C. Feasibility of Rate Vectors

We will define the feasibility of source rate vectors with respect to capacity functions. A capacity function simple assigns a capacity to each edge $e \in E$ that obeys node uplink constraints. A set of capacities $c(e)$ for each edge $e \in E$ is valid if for each node v, the sum of the capacities of the edges leaving v is at most $c_{out}(v)$, that is,

$$\sum_{e \in Out(v)} c(e) \leq c_{out}(v). \tag{1}$$

We denote $c : e \rightarrow [0, \infty)$ as the edge capacity function defined on E. Given upload capacities $c_{out}(v)$ for each node v in V, we say that a rate vector $z = \{z_s, s \in S\}$ is *achievable* if there exists an edge capacity function $c()$ satisfying (1) such that it is possible to broadcast $|S|$ independent sources of information at rates z_s from the each source $s \in S$ through the network to all receivers $r \in R_s$, possibly using network coding.

D. Our Contributions

An earlier result [6] established that routing is optimal for the single source multicast problem (with or without helpers) on a full-mesh P2P topology. First, we consider a simple extension of this result to the multi-source case and show that routing is optimal when coding is not allowed across sources. Second, and as the more important contribution, we show that for the multi-source case, when all groups have pairwise identical or disjoint receivers, routing is optimal even when coding across sources is allowed. Such a scenario is common in multi-party conferencing systems, hence our results have interesting practical applications in the design of infrastructure-less P2P multiparty conferencing systems (as described in the next section).

Given the hardness of the general problem for inter-session coding, we believe that the above result is both surprising and elegant. In contrast to the known results that inter-session coding is needed to achieve the maximum rate region in general topology, the unique structure of the P2P topology we consider in this paper allows us to achieve the maximum rate region for certain communication scenarios by packing only a linear number of trees per source.

E. Applications to P2P Multi-Party Conferencing

The problem we consider in this paper is motivated by and has applications to the design of P2P multi-party conferencing systems. Traditional multi-party conferencing (VoIP and/or video conferencing) is conducted using either a client-server architecture or in an ad hoc simulcast way. The client-server approach ensures that the entire upload bandwidth of each peer can be used for the delivery of just that peer's audio/video session; however, it places a heavy CPU and network bandwidth burden on the central server and thus incurs heavy deployment and egress ISP bandwidth costs. In the ad hoc simulcast approach, each user splits its uplink bandwidth equally among all receivers and sends its video to each receiver separately. Though simple to implement, this approach suffers from poor quality of service, especially when there is one peer with low upload bandwidth, as that peer is forced to use a low coding rate that degrades the overall experience of the other peers.

In contrast, a P2P approach for multiparty video conferencing *does not necessarily rely on centralized infrastructure and allows a peer to not only use its uplink to send its video stream but also to forward the video stream of other peers (with possibly lower uplink rates).* This approach facilitates efficient use of peer uplink bandwidth in the system and naturally accommodates peer uplink heterogeneity. Moreover, by accommodating helper nodes into our framework, our approach naturally allows other non-participant peers, ranging from infrastructure nodes (servers) to super nodes (peers with high capacity), to use their bandwidth to help further improving the quality of conferencing experience.

II. The Mutualcast Result and a Simple Extension to Multiple Sources

In the context of P2P topology with the above uplink constraint assumptions, a powerful theorem established in the Mutualcast paper [6] states the following. Consider a complete (directed) graph with node uplink constraints, consisting of a single source s, a set of receivers R_s (which includes s), and a set of helpers H_s. Then, the min-cut capacity for source s and receivers R_s can be achieved by packing at most $|R_s| + |H_s|$ Mutualcast trees as follows:

(1) One depth-1 tree rooted at s and reaching all receivers in $R_s - \{s\}$ over a single hop.

(2) $|R_s| - 1$ depth-2 trees, each rooted at s, reaching another receiver $r \in R_s - \{s\}$ over a first hop, and

then reaching all other receivers in $R_s - \{s, r\}$ over a second hop.

(3) $|H_s|$ depth-2 trees, each rooted at s, reaching a helper $h \in H_s$ over a first hop, and then reaching all receivers in $R_s - \{s\}$ over a second hop.

These trees are illustrated in Figure 1. This result extends and simplifies Edmonds' theorem [2] for a complete (directed) graph with node uplink constraints, in the sense that it allows helper (Steiner) nodes and uses only depth-1 and depth-2 Steiner trees.

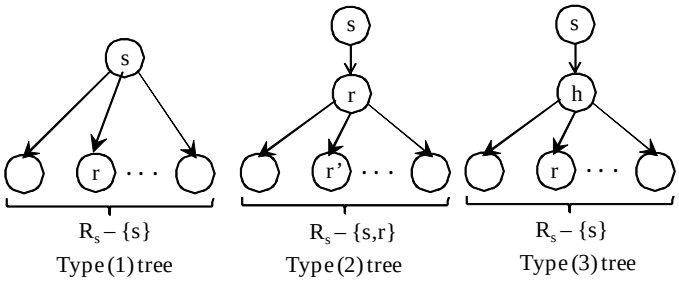

Fig. 1. Different types of Mutualcast trees.

Given that the Mutualcast Theorem is for single source multicast scenario only, we first extend this result to the case of multi-source multicast scenario when there is *no coding across sessions belonging to different sources*.

Theorem 1: For a complete (directed) communication graph $G = (V, E)$ with node uplink capacities, consider multiple multicast sessions given by source nodes $s \in S$, receiver set R_s, and helper nodes $H_s = V - R_s$ for session with source s. Then, the rate region $z = \{z_s, s \in S\}$ achievable by *network coding within each session* is also achievable by routing along (at most) $|R_s| + |H_s|$ Mutualcast trees for each source s independently.

Proof: Consider the realization of a given rate vector $z = \{z_s, s \in S\}$ through network coding within each session. Partition the capacity usage on each link according to portions used by each session. Then, by the Mutualcast theorem, the rate z_s for session with source s can be achieved by routing along $|R_s| + |H_s|$ Mutualcast trees using only the partitioned link capacities corresponding to this session. By superposing the Mutualcast trees used for each session, we get the claimed result. ∎

III. The Case of Inter-Session Network Coding with Identical or Disjoint Receiver Sets

In this section, we consider the scenario in which for any two sources s, s', the receiver sets $R_s, R_{s'}$ are either identical or disjoint. (Recall that a receiver set includes the source also.) Accordingly, we can partition the source and receiver nodes into disjoint subsets R^i, $i = 1, 2, \ldots, m$, for some $m \geq 1$, such that for each i and each source node in R^i, the set of receivers is R^i itself. (Note that each node in R^i need not be a source though.) Let the set of helper nodes be $H = V - \cup_i R^i$. The communication graph $G = (V, E)$ has *no edges between nodes in different R^i sets*. An edge between every other pair of nodes appears in G. That is, we do not allow two receivers to communicate directly unless they receive content from the same set of sources. This is illustrated in Figure 2. (Note that, as a special case, G is a complete graph when $m = 1$.) Let the j-th source in node set R^i be denoted by s^i_j and let its sending rate be z^i_j. For such a scenario, we show that routing is optimal and *inter-session network coding* is not needed. Moreover, routing along a linear number of Mutualcast trees per source is sufficient to achieve this.

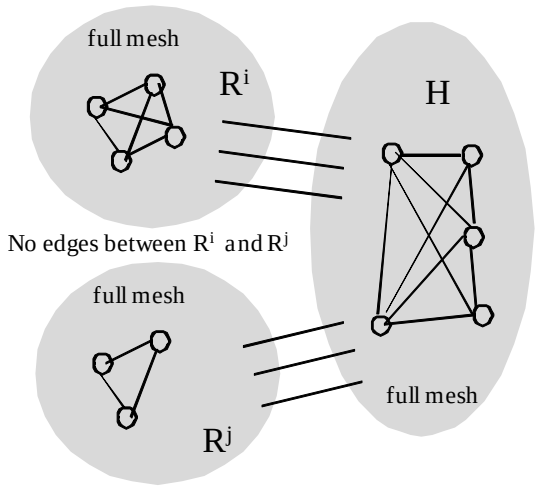

Fig. 2. Communication graph for the scenario in Theorem 2.

Theorem 2: For a (directed) communication graph $G = (V, E)$ with node uplink capacities, consider multiple multicast sessions given by source nodes $s \in S$ and receiver set R_s (including s itself) such for any two sources s, s', the receiver sets $R_s, R_{s'}$ are either equal or disjoint. That is, we can partition the source and receiver nodes into disjoint subsets R^i such that for each i and each source node in R^i, the set of receivers is R^i itself. Let the set of helper nodes be $H = V - \cup_i R^i$. The graph $G = (V, E)$ has all edges *except those between nodes in different R^i sets*. Then, the largest rate region $z = \{z^i_j\}$ achievable by *network coding across sessions* can also be achieved by routing along (at most) $|R_s| + |H|$ Mutualcast trees for each source s independently.

Proof: Let $z = \{z^i_j\}$ be an achievable rate vector. We will show by construction a set of Mutualcast trees achieving these rates.

Let $Out^i(h) = Out(h) \cap \cup_{r \in R^i} In(r)$ and $In^i(h) = In(h) \cap \cup_{r \in R^i} Out(r)$. That is, $Out^i(h)$ and $In^i(h)$ are respectively the outgoing and incoming edge sets at helper node h to and from nodes in R^i.

Consider an augmented graph $(\bar{V}, \bar{E})$ with $\bar{V} = V \cup \{s^i_0 : 1 \leq i \leq m\}$ and $\bar{E} = E \cup \{e^i_j\}$, where s^i_0 is a new "super-node" added to each receiver set R^i and e^i_j is a new edge connecting the super-node s^i_0 to source s^i_j in receiver set R^i with capacity $c(e^i_j) = z^i_j$, for $i = 1, \ldots, m$ and all respective j. Let $c(e)$ for the remaining edges $e \in E$ be given by an edge capacity function that achieves the rate vector z.

Suppose the super-node s^i_0 broadcasts information through the augmented graph $(\bar{V}, \bar{E})$ to the set of receivers R^i. Since $z = \{z^i_j\}$ is achievable in the original graph (V, E), it is

possible for super node s_0^i to broadcast information to R^i at the sum rate $\sum_j z_j^i$ in $(\bar{V}, \bar{E})$. However, an upper bound on the rate at which s_0^i can broadcast information to R^i is the minimum (over $r \in R^i$) of the value of the minimum cut (or equivalently the maximum flow) from s_0^i to r. Hence,

$$\sum_j z_j^i \leq \min_{r \in R^i} \text{mincut}(s_0^i, r).$$

Now for each $r \in R^i$, $\text{mincut}(s_0^i, r) \leq \sum_{e \in In(r)} c(e)$, since the sum of the capacities of the edges entering r is the value of a particular cut between s_0^i and r. Hence for each $r \in R^i$,

$$\sum_j z_j^i \leq \min_{r' \in R^i} \text{mincut}(s_0^i, r') \leq \sum_{e \in In(r)} c(e). \quad (2)$$

Thus, summing (2) over $r \in R^i$, we obtain

$$\begin{aligned}
|R^i| \sum_j z_j^i &\leq \sum_{r \in R^i} \sum_{e \in In(r)} c(e) \\
&= [\sum_{r \in R^i} \sum_{e \in In(r)} c(e) + \sum_{h \in H} \sum_{e \in In^i(h)} c(e)] - \sum_{h \in H} \sum_{e \in In^i(h)} c(e) \\
&= \sum_{v \in R^i \cup \{s_0^i\} \cup H} \sum_{e \in In(v)} c(e) - \sum_{h \in H} \sum_{e \in In^i(h)} c(e) \\
&= \sum_{v \in R^i \cup \{s_0^i\} \cup H} \sum_{e \in Out(v)} c(e) - \sum_{h \in H} \sum_{e \in In^i(h)} c(e) \quad (3) \\
&= \sum_{r \in R^i \cup \{s_0^i\}} \sum_{e \in Out(r)} c(e) + \sum_{h \in H} \sum_{e \in Out^i(h)} c(e) - \\
&\quad \sum_{h \in H} \sum_{e \in In^i(h)} c(e) \\
&= \sum_j z_j^i + \sum_{r \in R^i} \sum_{e \in Out(r)} c(e) + \sum_{h \in H} \sum_{e \in Out^i(h)} c(e) - \\
&\quad \sum_{h \in H} \sum_{e \in In^i(h)} c(e) \\
&\leq \sum_j z_j^i + \sum_{r \in R^i} c_{out}(r) + \sum_{h \in H} c_{out}^i(h) - \sum_{h \in H} \sum_{e \in In^i(h)} c(e) \quad (4)
\end{aligned}$$

where $c_{out}^i(h) = \sum_{e \in Out^i(h)} c(e)$ is the portion of $c_{out}(h)$ that is assigned to outgoing links from h to nodes in set R^i, for each $h \in H$, under capacity function $c()$. Note that (3) follows from the fact that in the induced (directed) subgraph on $R^i \cup \{s_0^i\} \cup H$, the sum of the capacities of the incoming edges is equal to the sum of the capacities of the outgoing edges, and (4) follows from (1). Hence, for all i, we have

$$\begin{aligned}
(|R^i| - 1) \sum_j z_j^i \quad \leq \quad & \sum_{r \in R^i} c_{out}(r) + \sum_{h \in H} c_{out}^i(h) - \\
& \sum_{h \in H} \sum_{e \in In^i(h)} c(e), \quad (5)
\end{aligned}$$

showing that $\sum_{r \in R^i} c_{out}(r)$ cannot be too small relative to z. In addition, we must have $z_s \leq c_{out}(s)$ for all $s \in S$, otherwise $z = \{z_j^i\}$ would not be achievable in (V, E).

We now come to the Mutualcast construction. For each *sender* $s \in S$, reserve bandwidth z_s from the upload capacity $c_{out}(s)$, and leave bandwidth $c_{out}(s) - z_s \geq 0$ unreserved. For each *non-sender* $v \in V - S$, leave the entire bandwidth $c_{out}(v)$ of the upload capacity unreserved. The *reserved* capacity for each sender $s \in S$ will be used to transmit at least one copy of the source out of the sender over the first hop of a Mutualcast tree. The *unreserved* capacity for each node $v \in V$ will be used for v to forward further copies of the source, on behalf of the sender if v is not itself the sender – if v is not a helper in H, it can forward copies of source s only if s is in the same receiver set R^i as v itself.

We are now going to greedily assign all the reserved and unreserved upload capacity to Mutualcast trees. For the Mutualcast construction, each source can use uplink capacities of its receivers and helpers in H only. For each sender $s \in S$ in turn, match every slice of bandwidth ϵ (for sufficiently small ϵ) of the sender's reserved capacity to an amount of unreserved capacity of any node v (in its receiver set or helper set H) with unreserved capacity that has not yet been matched to a sender. For matching up reserved capacity of a source with unreserved capacity, priority is given to first using uplink capacity of its receivers, and then to that of helpers in H.

Let source s belong to receiver set R^i. If v happens to be a helper, then the match causes a stream of bandwidth ϵ from s's *reserved* capacity to be routed to v, which relays the stream to the $(|R^i| - 1)$ receivers $r \in R^i - \{s\}$, thereby using up bandwidth $(|R^i| - 1)\epsilon$ of v's *unreserved* capacity. If v happens to be a receiver (other than the sender s), then the match causes a stream of bandwidth ϵ from s's *reserved* capacity to be routed to v, which relays the stream to the $(|R^i| - 2)$ receivers $r \in R^i - \{s, v\}$, thereby using up bandwidth $(|R^i| - 2)\epsilon$ of v's *unreserved* capacity. Finally, if v happens to be the sender s itself, then the match causes a stream of bandwidth ϵ from s's *reserved* capacity to be routed to a second receiver $r \in R^i - \{s\}$, and $(|R^i| - 2)$ streams of bandwidth ϵ from s's *unreserved* capacity to be routed to the $(|R^i| - 2)$ remaining receivers $r' \in R^i - \{s, r\}$, thereby using up bandwidth $(|R^i| - 2)\epsilon$ of s's unreserved capacity. These correspond to routings over Mutualcast trees of types (3), (2), and (1), respectively.

Thus, in total, assigning all $\sum_j z_j^i$ of the senders' rates in set R^i uses up at most $|R^i|E_3^i + (|R^i| - 1)E_2^i + (|R^i| - 1)E_1^i = (|R^i| - 1)(E_3^i + E_3^i + E_1^i) + E_3^i = (|R^i| - 1)\sum_j z_j^i + E_3^i$ of the total uplink capacity of nodes in R^i and H, where E_3^i, E_2^i, and E_1^i are the respective amounts of reserved bandwidth traversing Mutualcast trees of types (3), (2), and (1) for sources in set R^i. Since the construction gave priority to using uplink bandwidth of nodes in R_i over helper nodes in H, the uplink capacities of the former must be used up before that of the latter. Thus, we must have, for each i,

$$(|R^i| - 1) \sum_j z_j^i + E_3^i \leq \sum_{r \in R^i} c_{out}(r) + \sum_{h \in H} c_{out}^i(h) \quad (6)$$

We shall show that, for each i,

$$E_3^i \leq \sum_{h \in H} \sum_{e \in In^i(h)} c(e), \quad (7)$$

whence we can conclude that assigning all $\sum_j z_j^i$ of the senders' rates in R^i uses up at most

$$(|R^i| - 1) \sum_j z_j^i + \sum_{h \in H} \sum_{e \in In^i(h)} c(e)$$

of the uplink capacities of nodes in R^i and helper nodes in H, which is at most the total uplink capacity $\sum_{r\in R^i} c_{out}(r) + \sum_{h\in H} c^i_{out}(h)$, according to (5). Hence the greedy construction of Mutualcast trees can always be accomplished, with information routed along the resulting trees at rates z^i_j.

From the Mutualcast construction, it is clear that $E^i_3 \leq \frac{1}{|R^i|-1} \leq \sum_{h\in H} c^i_{out}(e)$. Then (7) follows from the fact that, for each i,

$$\sum_{h\in H} c^i_{out}(h) \leq (|R^i|-1) \sum_{h\in H} \sum_{e\in In^i(h)} c(e) \qquad (8)$$

Hence, we need to show that (8) holds for at least one valid capacity function that can achieve the rates $z = \{z^i_j\}$ under coding.

If this is not true for the capacity function $c()$, we will construct another valid capacity function $c'()$ which obeys (8) and achieves the same rates z^i_j. Assume, then, that for some i,

$$\sum_{h\in H} c^i_{out}(h) > (|R^i|-1) \sum_{h\in H} \sum_{e\in In^i(h)} c(e). \qquad (9)$$

This allows us to modify the coding solution that achieves the rates z^i_j as follows. We will use the helpers to simply replicate any bits it receives from a node in R^i to all other nodes in R^i under the given bound $c^i_{out}(h)$ on the portion of uplink capacity of each helper node h that is assigned to outgoing links in R^i. Clearly, this is possible since, according to (9), the total uplink capacity of helpers to nodes in R^i is more than $(|R^i|-1)$ times the total raw bits entering each helper from nodes in R^i. In particular, if node $j \in R^i$ was sending flow of rate x_j in aggregate to all nodes $h \in H$ under the coding solution, it will now send out a fraction

$$\frac{c^i_{out}(h)}{\sum_{h\in H} c^i_{out}(h)}$$

of x_j to helper node h. Then, if each helper h replicates the bits it receives from any $j \in R$ to all other nodes in R^i, we can verify, using (9) that this obeys uplink constraints at all helper nodes $h \in H$. This is illustrated in Figure 3.

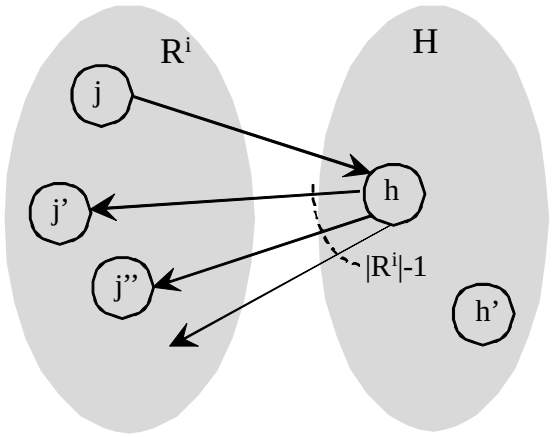

Fig. 3. Modifying coding solution using replication at each $h \in H$ to all $j \in R^i$.

Hence, all the information entering H from R^i can now be broadcast to every node in R^i. Since each receiver $r \in R^i$ was earlier able to decode the bits it is supposed to receive from source nodes in R^i *regardless of the bit values sent by source nodes outside* R^i, the coding that may have been done in H earlier can thus be done at each node $j \in R^i$ by *assuming any fixed value (say '1') for bits that come from source nodes outside* R^i. This guarantees that, under the modified coding scheme, each receiver $r \in R^i$ can still decode the bits it was supposed to receive from source nodes in R^i.

On the other hand, all nodes (sources, receivers, or helpers) outside R^i can still perform the coding and receivers still decode bits from their respective sources, by *assuming any fixed value (say '1') for bits that come from source nodes within* R^i. Therefore, the rates z^i_j are still achieved after the modification.

The capacity usage on edges between R^i and H (for all such i for which (9) holds) is set as per usage by the above modification. The capacity usage on other edges remains the same. Call this new capacity function $c'()$. By the modification, it follows that, for all such i,

$$c'^i_{out}(h) = (|R^i|-1) \sum_{e\in In^i(h)} c'(e) \quad \text{for all } h \in H \qquad (10)$$

By summing (10) over all $h \in H$, it follows that the modified capacity function $c'()$ obeys (8) (with equality) for all such i. (It continues to hold for all other i.) It also achieves the same rates z^i_j. This completes the proof. ■

IV. Conclusion and Open Problems

We have established the optimality of routing for certain multi-source multicast communication scenarios with node uplink constraints. The assumption of full connectivity (complete communication graph) could be impractical to realize in practice for large-scale P2P applications, where peer nodes are limited by resources on the number of other peers that they can maintain direct connections to. The extension of the result to (or, a counter-example for) arbitrary receiver sets and the investigation of achievable rate region and optimality of routing (vs. network coding) under degree constraints in the overlay are challenging open problems.

References

[1] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung. Network information flow. (4):1204–1216, July 2000.

[2] J. Edmonds. Edge-disjoint branchings. *Combinatorial Algorithms, R. Rustin, ed.*, pages 91–96, 1973.

[3] H. N. Gabow and K. S. Manu. Packing algorithms for arborescences (and spanning trees) in capacitated graphs. *Mathematical Programming*, 82(1-2):83–109, June 1998.

[4] S. Jaggi, P. Sanders, P. A. Chou, M. Effros, S. Egner, K. Jain, and L. Tolhiuzen. Polynomial time algorithms for multicast network code construction. *IEEE Transactions on Information Theory*, 51(6):1973–1982, 2005.

[5] K. Jain, M. Mahdian, and M. R. Salavatipour. Packing steiner trees. In *14th ACM-SIAM Symp. on Discrete Algorithms (SODA)*, Jan. 2003.

[6] J. Li, P. A. Chou, and C. Zhang. Mutualcast: an efficient mechanism for content distribution in a p2p network. In *Proceedings of Acm Sigcomm Asia Workshop*, Beijing, China, Apr. 2005.

[7] X. Yan, R. W. Yeung, and Z. Zhang. The capacity region for multi-source multi-sink network coding. In *2007 IEEE International Symposium on Information Theory (ISIT 2007)*, Nice, France, June 2007.