On Robust Network Coding Subgraph Construction under Uncertainty

Christopher S. Chang, Tracey Ho, and Michelle Effros
Department of Electrical Engineering
California Institute of Technology
Pasadena, CA 91125, USA
E-mail: {cswchang, tho, effros}@caltech.edu

Abstract—We consider the problem of network coding subgraph construction in networks where there is uncertainty about link loss rates. For a given set of scenarios specified by an uncertainty set of link loss rates, we provide a robust optimization-based formulation to construct a single subgraph that would work relatively well across all scenarios. We show that this problem is coNP-hard in general for both objectives: minimizing cost of subgraph construction and maximizing throughput given a cost constraint. To solve the problem tractably, we approximate the problem by introducing path constraints, which result in polynomial time-solvable solution in terms of the problem size. The simulation results show that the robust optimization solution is better and more stable than the deterministic solution in terms of worst-case performance. From these results, we compare the tractability of robust network design problems with different uncertain network components and different problem formulations.

I. INTRODUCTION

Network coding has recently been shown to improve performance over both wired and wireless communication networks [1], [2]. In multicast communications, previous research has shown that the network coding can achieve the maximum throughput of a given network that is equal to the maximum flow between the source and each destination. In addition to network throughput benefits, network coding can provide robustness to uncertain communication links and uncertain network topologies [3]. Accordingly, throughput gain as well as robustness are well-known advantages of network coding.

In this paper, we consider the problem of network coding subgraph construction that is robust against uncertainty about link loss rates. For a given set of scenarios specified by an uncertainty set of link loss rates, the goal is to construct a single subgraph that works relatively well across all scenarios. To achieve the goal, an optimization problem with unknown variables is considered. In such optimization problems containing unknown variables, the most developed approaches are worst-case analysis and stochastic optimization. But scenario-based stochastic optimization cannot be used unless the uncertainty is probabilistic. On the other hand, worst-case analysis sometimes results in more conservative solutions. In this paper, we follow the robust optimization approach of optimizing the worst-case performance introduced in [4].

Recently, robust optimization research has concentrated on robust convex optimization (including linear optimization). Also, the results on robust optimization have been applied to a number of network optimization problems. Applegate et al. suggested a robust routing which guarantees a nearly optimal utilization of a network against uncertain traffic demands [5]. Mudchanatongsuk et al. showed that the network optimization problem with demand uncertainty can be solved in polynomial time if additional path constraints are imposed [6]. Ordóñez et al. provided conditions for demand and cost uncertainty sets to make the network optimization problem tractable [7]. In contrast to those polynomial time-solvable problems, Atamtürk et al. showed that a two-stage robust optimization for a multi-commodity network flow and design problem with discrete design variables under demand uncertainty is NP-hard [8]. Chekuri et al. proved coNP-hardness of the single-source robust network design problem under demand uncertainty [9], which we state precisely in Section III.

To the best of our knowledge, most of the robust optimization problems in the literature of networking have considered the uncertainty about demand or link cost. In contrast, we focus on the problem under the uncertainty about link status—link loss rates (or interchangeably, link success probabilities) —in this paper. The contribution of this paper is as follows: we prove the coNP-hardness of the problem for objectives of minimizing cost and maximizing throughput; on the other hand, we provide a polynomial-time solvable problem formulation by introducing path constraints that approximate the problem.

The structure of the paper is as follows: In Section II, we describe our network model and problem formulation. In Section III, we consider the tractability of the problem: we prove the general problem is coNP-hard for min-cost objective and max-throughput objective. In Section IV we show that the problem with path constraints is polynomial time-solvable in terms of problem size. In Section V, we provide simulation results for simple network examples in which we compare the robust optimization solution with other method. Finally, Section VI concludes the paper and contains some remarks with respect to future directions.
We consider a directed graph $G = (\mathcal{N}, \mathcal{E})$ where $\mathcal{N}$ is the set of nodes and $\mathcal{E}$ is the set of links (arcs), $m = |\mathcal{N}|$ is the number of nodes and $n = |\mathcal{E}|$ is the number of links. To simplify the problem, we consider a network coding problem with a single source node and a single sink node, where network coding provides a robustness benefit.

For performance metrics, we have two different objective functions: minimizing cost of subgraph construction and maximizing throughput for a given cost constraint. For the former problem, we want to find a min-cost network coding subgraph that satisfies a demand requirement for the sink node. For the latter problem, we want to find a max-throughput network coding subgraph that satisfies a cost constraint on the chosen subgraph.

We denote by $P_{ij}$ the set of simple paths from node $i$ to node $j$ ($P := \bigcup_{ij} P_{ij}$), $c$ and $\text{cap}$ are a given path cost vector and a given link capacity vector, respectively. $bgt$ is a budget for subgraph construction when the objective function is the throughput of network, and $D$ is the demand required for the sink node when the objective function is the cost of subgraph construction. We introduce a budget, $bgt$, and a demand, $D$, to avoid trivial solutions such as flooding and zero flow in each case, respectively.

We have the following uncertainty and decision variables:

- Uncertainty: $w$ (vector of path success probabilities) taking values in an uncertainty set $W$

- Optimization variables:
  - $k$ (network coding subgraph; $k_p$ is the maximum feasible flow on path $p \in P$) determined prior to the realization of the path success probabilities.
  - $h$ (vector of actual path flows; $h_p$ is the flow along the subgraph, $k_p$, in the presence of path loss rates, i.e. $h_p \leq k_p w_p$, $\forall p \in P$) determined after the realization of the path success probabilities.

Thus, we have a two-stage optimization problem (called the Adjustable Robust Counterpart (ARC) problem [7], [10]). This formulation fully captures the robustness properties of network coding. Since the actual flow is determined after the realization of path success rates, network coding intrinsically exploits the best routes for the actual flow within the capacity constraints from the predetermined subgraph. In contrast, a single-stage robust optimization (called the Robust Counterpart (RC) problem [4]) would require fixing a flow $h$ feasible under any realization of the link qualities, resulting in a much more conservative solution. Since ARC has a larger robust feasible set, the solution for ARC is at least as good as the one for RC. However, a two-stage optimization increases the problem complexity significantly. It is known that the ARC problem of a linear program with polyhedral uncertainty set is NP-hard in general [11]. We describe the tractability of this problem in more depth in the next section.

### III. HARDNESS OF ROBUST SUBGRAPH SELECTION

#### A. Min-Cost Criterion

The robust minimum cost subgraph selection problem with the demand requirement in terms of path flows is as follows:

$$\begin{align*}
\min_k & \quad \sum_{p \in P} c_p k_p \\
\text{s.t.} & \quad \sum_{p \in P : l \in p} k_p \leq \text{cap}(l), \forall l \in \mathcal{E} \\
& \quad \text{for all } w \in W, \exists h : \\
& \quad \{ h_p \leq k_p w_p, \forall p \in P \} \\
& \quad \{ \sum_{p \in P} h_p \geq D \}
\end{align*}$$

Note that any network flow problem can be formulated using path and cycle flows and vice versa [12, Theorem 3.5]. Thus, we can find the corresponding link formulation as follows. If $d$ is a scalar for the demand, then we can write a supply-demand vector as $d(e_s - e_t)$, where $e_s, e_t$ are canonical vectors for source and sink, respectively. Let $N$ denote a node-arc incidence matrix [12]. Note that any variables with tilde are arc-flow variables and $\tilde{W}$ is the uncertainty set for the link success probabilities. Then, the corresponding link formulation is as follows:

$$\begin{align*}
\min_k & \quad \sum_{l \in \mathcal{E}} \tilde{c}(l)\tilde{k}(l) \\
\text{s.t.} & \quad 0 \leq \tilde{k}(l) \leq \text{cap}(l), \forall l \in \mathcal{E} \\
& \quad \text{for all } \tilde{w} \in \tilde{W}, \exists \tilde{h}, d : \\
& \quad \{ \tilde{h}(l) \leq \tilde{k}(l)\tilde{w}(l), \forall l \in \mathcal{E} \} \\
& \quad \{ N \cdot \tilde{h} = d(e_s - e_t) \} \\
& \quad d \geq D
\end{align*}$$

We consider hardness of (1) and (2) under polyhedral uncertainty sets for path success probabilities $W$ and link success probabilities $\tilde{W}$, respectively. We show the complexity of this problem by reduction from the single-source robust network design problem under demand uncertainty, which is known to be coNP-hard for undirected and directed graphs [9]. An instance of the latter is defined by a given graph $G = (\mathcal{N}, \mathcal{E})$, a link cost vector $c$, a single source node $s \in \mathcal{N}$ and a convex polyhedral set $D$ of demand matrices such that for each $D \in D$, the demanded flow $D_{ij}$ from node $i$ to node $j$ is zero for non-source nodes $i \neq s$. The objective is to find the least cost vector of link capacity reservations $u$ sufficient to support a multi-commodity fractional routing for each demand matrix in $D$, i.e.

$$\begin{align*}
\min_u & \quad \sum_l c(l)u(l) \\
\text{s.t.} & \quad \text{for all } D \in D, \exists f \text{ satisfying} \\
& \quad \{ \sum_{p \in P_{ij}} f_p = D_{ij}, \forall i, j \in \mathcal{N} \} \\
& \quad \{ \sum_{ij} \sum_{l \in \mathcal{E}} f_p \leq u(l), \forall l \in \mathcal{E} \}
\end{align*}$$

where $f$ is a vector specifying the flow $f_p$ on each path $p$. Note that the routing may change for different demand matrices—thus, it is not a path-constrained problem. We will show that an instance of our problem (2) is equivalent to Fig. 1 (a).
Theorem 1. The robust minimum cost subgraph selection problems (1) and (2) are coNP-hard for polyhedral uncertainty sets $W$ and $\tilde{W}$, respectively.

**Proof:** Note that for the robust network design problem (3), if each demand matrix in $D$ is scaled by a constant factor, then the optimal link capacity reservation vector $u$ is scaled by the same factor. Thus without loss of generality, we consider an instance $H$ of problem (3) with graph $G = (\mathcal{N}, \mathcal{E})$, source $s$ and convex polyhedral set $D$ of demand matrices such that $\sum_{t} D_{ij} \leq 1 \forall D \in D$, illustrated in Fig. 1 (a).

From this we obtain an instance $H'$ of the robust minimum cost subgraph selection problem (2), illustrated in Fig. 1 (b), as follows. We add to network $G$ an additional node $t'$; $s$ and $t$ are the source and sink nodes respectively. We introduce an additional link $l'_i$ of capacity 1, cost 0 and success probability $w_{it'}$ from each node $i \in \mathcal{N}$ to $t'$. The vector $w$ of success probabilities $w_{it}$ lies in the uncertainty set

$$w_{it} = \frac{1 - \sum_{j} D_{ij}}{D} \quad \forall i \neq s \quad \forall D \in D$$

which is a convex polyhedral set. The links in $\mathcal{E}$ have success probability 1 and no capacity constraints. The demanded $s-t$ flow is 1. This requires the solution to have $\tilde{k}(l'_i) = 1$ and, for each $D \in D$, there must exist a multicommodity flow $\tilde{f}$ of size $D_{si}$ from $s$ to each other node $i \in \mathcal{N}$ satisfying $\tilde{h}(l'_i) \leq \tilde{k}(l'_i).$ Thus, an optimal solution for the robust minimum cost subgraph selection problem also solves the single-source robust network design problem. Note that since no $s-t$ path contains more than one of the uncertain links, (4) corresponds to a polyhedral uncertainty set in terms of path success probabilities: all paths through link $l'_i$ ($i \neq s$, $i \in \mathcal{N}$) have path success probability $D_{si}$, and the path through the link $(s, t)$ has path success probability $1 - \sum_{j} D_{ij}$. Then by using the same proof, the path formulation (1) is also coNP-hard. This completes the proof.

Fig. 1. Two equivalent network optimization problems with different uncertainties: (a) Minimizing cost under demand uncertainty (b) Minimizing cost under uncertainty of link success probabilities

B. Max-Throughput Criterion

The path formulation for the robust maximum throughput subgraph selection problem with the cost constraint is as follows:

$$\max_{k,h} \sum_{p \in P} h_p$$

$$\text{s.t.} \quad \sum_{p \in \mathcal{P} : l \in \mathcal{E}} k_p \leq \text{cap}(l), \forall l \in \mathcal{E}$$

$$c^T \cdot k \leq bgt$$

$$\text{for all } w \in W, \exists h :$$

$$h_p \leq k_p w_p, \forall p \in \mathcal{P}$$

Similarly to the min-cost problem, we can formulate the corresponding link formulation as follows:

$$\max_{k} d$$

$$\text{s.t.} \quad 0 \leq \tilde{k}(l) \leq \text{cap}(l), \forall l \in \mathcal{E}$$

$$\sum_{l} \tilde{c}(l) \tilde{k}(l) \leq bgt$$

$$\text{for all } w \in \tilde{W}, \exists h, d :$$

$$\tilde{h}(l) \leq \tilde{k}(l) \tilde{w}(l), \forall l \in \mathcal{E}$$

$$\mathbf{N} \cdot \tilde{h} = d(\mathbf{e}_s - \mathbf{e}_t)$$

The link formulation corresponding to (6) can be found similarly. Now, we prove coNP-hardness of the max-flow subgraph selection problem.

Theorem 2. The robust maximum flow subgraph selection problems (5) and (6) are coNP-hard for polyhedral uncertainty sets $W$ and $\tilde{W}$, respectively.

**Proof:** We consider an instance $H$ of problem (3) with graph $G = (\mathcal{N}, \mathcal{E})$, source $s$ and convex polyhedral set $D$ of demand matrices such that $\sum_{t} D_{ij} \leq 1 \forall D \in D$, and construct the instance $H'$ of problem (2) exactly as in the proof of Theorem 1. Let $C$ be the optimal cost of $H$.

2213
From $H'$ we construct an instance of problem (6) as follows. We add another link $l'$ from $s$ to $t$ that has cost $c' > n \max_{l \in E} c(l)$, success probability 1 and no capacity constraint, and set $bgt > C$. The optimal solution has $k(l') = (bgt - C)/c'$, and contains a solution for $H$. Similarly to the min-cost problem, the max-flow problem (5) with polyhedral uncertainty for path success probabilities can also be shown to be coNP-hard. This completes the proof.

IV. PATH FORMULATION

Instead, if we introduce path constraints to the network optimization problem, the problem becomes polynomial time-solvable. In this section, we describe the approximate solution for the max-throughput problem only. The approximate solution for the min-cost problem is exactly analogous.

We introduce a path constraint, $h_p = k_p w_p$, to replace (1) in (5): the actual flow is an affine function of the uncertainty (called the Affinely Adjustable Robust Counterpart (AARC) [10]). Thus, it becomes an approximate problem. Now we can replace $\sum_{p \in P} h_p$ with $\sum_{p \in P} k_p w_p$ and formulate $\max_k - \min_w$ problem as follows:

$$\max_k \min_w \sum_{p \in P} k_p w_p$$

$$\text{s.t.} \sum_{p \in P \setminus \{l\}} k_p \leq \text{cap}(l), \forall l \in \mathcal{E}$$

$$c^T \cdot k \leq bgt$$

$$w \in W$$

(7)

By using duality of the minimization problem, we can combine the minimization problem with the maximization problem. If $W$ is a convex polyhedron set, then the combined problem becomes a single linear program (LP). Let $z_{\text{ROS}}$ and $k_{\text{ROS}}$ denote the optimum objective value (Max-throughput) and the optimum subgraph, respectively.

$$z_{\text{ROS}} = \max_{k,r,\lambda,\mu} \sum_{l \in \mathcal{E}} k_p \leq \text{cap}(l), \forall l \in \mathcal{E}$$

$$\text{s.t.} c^T \cdot k \leq bgt$$

$$-\lambda T \cdot g - \mu^T \cdot g_{\text{eq}} \geq r$$

$$k + H^T \cdot \lambda + H_{\text{eq}}^T \cdot \mu \geq 0$$

$$k, \lambda \geq 0, \quad r \geq 0$$

(8)

Therefore, for a case with tractable problem size where we have a polyhedral uncertainty set for path success probabilities, we can solve this problem by using any efficient LP solving algorithm such as Interior-Point method [13]. Although we have a single LP formulation, however, the problem size, particularly the number of paths, grows exponentially in the size of the network in general. Therefore, for a large network, LP formulation (8) can be intractable due to the enormous problem size. In such a case, we speculate that we may use the column generation approach [12], but we have not proved its application.

V. EVALUATIONS FOR PATH FORMULATION

In this section, we evaluate the performance of the path formulation in four simple network examples shown in Fig. 3. We compare the two-stage robust optimization with the non-robust optimization in terms of the worst-case performance. For simplicity, we consider a box uncertainty set with a linear equality constraint:

$$W = \{ w : a_i \leq w_i \leq b_i, \forall i \in P, H_{\text{eq}} \cdot w = g_{\text{eq}} \}$$

Each path has a nominal path success probability, $\pi_i = (a_i + b_i)/2$, with variations around the nominal value; in addition, the equality constraint, $H_{\text{eq}} \cdot w = g_{\text{eq}}$, satisfied by nominal values as well, is introduced to avoid a trivial solution. Furthermore, the equality constraint normalizes the max-flow over the uncertain scenarios.

A. Deterministic Case with Nominal Values

First, we consider the deterministic case with nominal values for uncertain path success probabilities, $w$. Let $z_{\text{DTM}}$ and $k_{\text{DTM}}$ denote the optimum objective value (max-throughput) and the optimum subgraph, $k$, respectively.

$$z_{\text{DTM}} = \max_k \sum_{p \in P} k_p \pi_p$$

$$\text{s.t.} \sum_{p \in P \setminus \{l\}} k_p \leq \text{cap}(l), \forall l \in \mathcal{E}$$

$$c^T \cdot k \leq I$$

(9)

Note that this method corresponds to a single-stage non-robust optimization strategy.

B. Maximum Flow with Given Subgraph

To make the problem (5) tractable, we have assumed that the second stage variables, $h$, are affine functions of the uncertainty, (i.e. $h_p = w_p h_p$), but in reality, the second stage variables can indeed change arbitrarily. Therefore, we can evaluate the max-flow for a given subgraph with arbitrarily chosen second stage variables. To evaluate the approximate solution, we compare the worst-case performances (max-flow) for given subgraphs. We can formulate an optimization problem for the worst-case performance with a new capacity bound, $u_l(S, w)$, the usage of link $l$ of a subgraph $S$ as follows:

$$z_{\text{WC}}(S) = \min_{w \in W} \max_k \sum_{p \in P} h_p$$

$$\text{s.t.} \sum_{p \in P \setminus \{l\}} h_p \leq u_l(S, w), \forall l \in \mathcal{E}$$

(10)
TABLE I
FREQUENCY WHEN \( k_{\text{ROS}} \) PROVIDES BETTER WORST-CASE PERFORMANCE THAN \( k_{\text{DTM}} \) (TOTAL 200 RANDOM TRIALS)

<table>
<thead>
<tr>
<th>Network</th>
<th>( z_{W,C}(k_{\text{ROS}}) \geq z_{W,C}(k_{\text{DTM}}) )</th>
<th>Avr. Perf. Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>95%</td>
<td>12.3%</td>
</tr>
<tr>
<td>2</td>
<td>97%</td>
<td>8.7%</td>
</tr>
<tr>
<td>3</td>
<td>97.5%</td>
<td>38.3%</td>
</tr>
<tr>
<td>4</td>
<td>98%</td>
<td>11.8%</td>
</tr>
</tbody>
</table>

where \( u_i(S, w) \) can be found as follows:

\[
u_i(k_{\text{ROS}}, w) \equiv \sum_{p \in P : i \in P} k_{\text{ROS}} p w_p \quad \text{from (8)}
\]

\[
u_i(k_{\text{DTM}}, w) \equiv \sum_{p \in P : i \in P} k_{\text{DTM}} p w_p \quad \text{from (9)}
\]

However, when we reduce (10) into one single optimization problem, the inequality constraint results in minimizing a non-convex quadratic function, which is NP-hard in general. Therefore, instead, we generate 5,000 random samples for path success probabilities uniformly from a given uncertainty set \( W \) and compare the minimum values among them. Using this random simulation, we compare the robust optimization solution and the deterministic solution in terms of the worst-case performance in the network examples, shown in Fig. 3.

**C. Results**

From the simulation, we observe that the robust optimization solution results in better performance than the non-robust optimization in terms of worst-case performance over 90% of times. The performance gain of the robust optimization solution against the non-robust strategy is around 10% for all examples. In addition, the robust optimization solution provides more stable outputs whereas the deterministic solution sometimes results in poor performance.

Recall that the path constraint that is introduced to make the problem tractable results in approximation of the original problem. So, the worst-case max-flow for a robust optimization solution without path constraints is always at least as good as the solution with path constraints, i.e. \( z_{W,C}(k_{\text{ROS}}) \geq z_{\text{ROS}} \) always holds.

The simulation results are summarized in Table I and Fig. 4.

**VI. CONCLUSIONS AND FUTURE WORK**

We have described the problem of network coding subgraph construction in networks where there is uncertainty about link success probabilities. We formulated the problem using the best worst-case guaranteed robust optimization technique. However, we proved that the problem is coNP-hard for the min-cost objective and the max-throughput objective. Accordingly, we suggested a tractable approximate solution by using path constraints. The tractability of network optimization problems with different problem formulations and different uncertainties is summarized in Table II.

We see many directions for further research. First, we can perform complexity analysis for the approximate solution of the two-stage optimization. Second, we can apply the results to more general network problems, such as multicast and multicommodity flows. Lastly, it is still an interesting problem to seek a new approximate algorithm to solve the problem in polynomial time.

**REFERENCES**


