Approximation Algorithms for Fragmenting a Graph against a Stochastically-Located Threat

David B. Shmoys[1,*] and Gwen Spencer[2,**]

[1] School of ORIE and Dept. of Computer Science, Cornell University, Ithaca, NY
shmoys@cs.cornell.edu
[2] School of ORIE, Cornell University, Ithaca, NY
gms39@cornell.edu

Abstract. Motivated by issues in allocating limited preventative resources to protect a landscape against the spread of a wildfire from a stochastic ignition point, we give approximation algorithms for a new family of stochastic optimization problems.

1 Introduction

Increasing frequency of catastrophically-damaging wildfire events has stimulated interest among foresters and land managers in effective use of preventative fuel reductions. Traditional fire suppression policy has focused almost exclusively on realtime firefighting (once the fire has broken out), but preventative fuel reductions such as dead-brush removal, small-scale controlled burns, and crown raising can be applied in advance to slow or stop the spread of wildfires. Recent wildfire modeling literature has used historical and scientific information to estimate a distribution of wildfire occurrence in which both the ignition site and the wind direction can vary [3],[8].

The planning problem of how to allocate limited resources across preventative and realtime stages, and where to distribute preventative resources using probabilistic information motivates a natural new family of budgeted stochastic optimization problems that fragment (or cut) a landscape graph to isolate a stochastically occurring ignition point. A key feature is the tradeoff between spending preventively when only distributional knowledge is available and spending at increased cost once a fire has broken out. We explore a number of model variants. Studying this family of problems through the lens of efficient approximation, we give constant bicriteria approximations in trees, and a budget-balanced constant approximation for the limiting case in which real-time actions become prohibitively expensive. Our techniques also yield new approximation results for multistage stochastic extensions of the budgeted Maximum Coverage problem. The theme of our models (protecting a network from the spread of a stochastic

* Work supported under grants no. CCR-0635121, DMS-0732196, CCF-0832782, CCF-1017688.

** Supported in part by NSF under a Graduate Research Fellowship and under grants no. CCR-0635121, DMS-0732196, CCF-0832782, CCF-1017688.

R. Solis-Oba and G. Persiano (Eds.): WAOA 2011, LNCS 7164, pp. 123–136, 2012.

outbreak of a harmful diffusive process) has other important environmental applications (e.g., containing invasive species over land or through water systems).

Results. In trees, the problem is (weakly) NP hard even when there is a single ignition point that is known deterministically [4] (the Knapsack Problem is a special case). An existing PTAS in graphs of bounded treewidth for the deterministic ignition-point case extends immediately to a PTAS in graphs of bounded treewidth for the deterministic ignition-set case. Applying some careful partial-enumeration then allows a PTAS in trees for the stochastic case in which the number of scenarios is constant.

The Graph Protection Problem: Summary of Main Results		
	restricted graph classes	general graphs (via Räcke)[5]
2-stage		
stochastic, single source	trees: $(1-(1-1/2\delta)^{2\delta}, 2)$ Via pipage rounding. *Alternative:* (0.387, 1)	constant number of scenarios $\Rightarrow$ $(1-(1-1/2n)^{2n}, O(\log n))$
stochastic, single source with (B_1, B_2)	trees: $(1-(1-1/2\delta)^{2\delta}, 1, 2)$ Via pipage rounding.	constant number of scenarios $\Rightarrow$ $(1-(1-1/2n)^{2n}, O(\log n), O(\log n))$
k-stage		
stochastic, single source	trees, restricted partition hierarchy: $(1-(1-1/k\delta)^{k\delta}, 2+\epsilon)$ Via pipage rounding.	constant number of scenarios and restricted partition hierarchy $\Rightarrow$ $(1-(1-1/kn)^{kn}, O(\log n))$
1-stage		
stochastic, single source with probabilistic edges	trees: $(1-1/e, 1)$ Due to submodularity.	open
stochastic, single source	trees: $(1-(1-1/\delta)^{\delta}, 1)$ Reduce to MCKP, apply [1].	$(1-(1-1/n)^{n}, O(\log n))$
stochastic with constant support and constant source size	trees: $(1+\epsilon, 1)$	$(1+\epsilon, O(\log n))$
deterministic with arbitrary source size	bounded tree width: $(1+\epsilon, 1)$	$(1+\epsilon, O(\log n))$
deterministic with single source	bounded tree width: $(1+\epsilon, 1)$ [4]	$(1+\epsilon, O(\log n))$ [4]

For the 2-stage stochastic model in which actions may either be taken in advance of the ignition based on probabilistic information, or after the single ignition point is known at inflated cost, we give a $(1-(1-1/2\delta)^{2\delta})$-approximation in trees which violates the budget by a factor of at most 2 (δ is the tree diameter). Notably, the inflation in the second stage can vary across scenarios and edges. For the limiting stochastic case in which no realtime action is possible, we give a $(1-(1-1/\delta)^{\delta})$-approximation algorithm in trees for the case of probabilistic ignition from a single source. We also give a 0.387-approximation which is budget-balanced for the 2-stage stochastic model, and some results for a k-stage extension. In some cases we can extend to general graphs with an additional $O(\log n)$ loss in budget-balancedness via the probabilistic cut-capacity approximation result of Räcke [5] as in Engelberg, et al. [2].

For an extension in which transmission on edges is probabilistic and depends on the level of investment in removing the edge (assuming independence of edge realizations), we give a $(1 - 1/e)$-approximation algorithm in trees.

Our multistage and probabilistic-transmission results in trees also hold for analogous generalizations of the Maximum Coverage with Knapsack Constraint problem (MCKP) in which elements may *fail* independently with probability that depends on the level at which we invest in them, and the objective is to maximize the expected weight of the sets covered by the realized elements. For probabilistic element-failure MCKP, our guarantee matches the asymptotic guarantee for the deterministic element case from Ageev & Sviridenko [1].

Related Literature. The placement of preventative fuel treatments has been addressed in the recent forestry literature. Finney [3] prioritizes spatial fire spread dynamics, limits probabilistic model components, and aims to reduce the rate of spread of the head of fire. Wei et al. [8] considers the objective of reducing expected value lost across a grid-cell landscape by reducing burn probabilities (probabilities computed through simulation); however their IP-based approach is based on a questionable linearity assumption. These approaches produce divergent solution forms: the development of additional mathematical tools and techniques that simultaneously address stochastic and spatial aspects would be useful to decision-makers faced with this important planning problem.

The problems we study have ties to the existing computer science literature. The special case in which the ignition point is known deterministically and there is a single decision stage has been studied as the Minimum-Size Bounded-Capacity Cut problem by Hayrapetyan et al. [4]. They show that the problem is weakly NP-hard in trees by reduction from the Knapsack problem. In general graphs they give two different $(\frac{1}{1-\lambda}, \frac{1}{\lambda})$ bicriteria-approximations for the (expected value burned, budget), and they give a PTAS in graphs of bounded tree width. Engelberg, et al. [2] study a number of budgeted cut problems in graphs including the weighted Budgeted Separating Multiway Cut Problem (wBSMC), which the single-stage (aka, no realtime action) stochastic version of our problem reduces to. They apply Räcke's probabilistic cut-capacity-preserving approximation to reduce to the case of trees, then observe submodularity in trees, and apply [7] to get a $((1 - 1/e),\ O(\log n))$ bicriteria result. Our LP-based result for the single-stage stochastic version of our problem in trees generalizes to wBSMC in trees giving a slightly stronger $(1 - (1 - 1/n)^n,\ O(\log n))$ bicriteria result.

Techniques. For the deterministic case, a psuedopolynomial-time exact dynamic programming method is converted to an efficient scheme by rounding the input (as in [4]): our extension to general ignition sets is by demonstrating bounded treewidth of a modified input. For the extension with probabilistic-edge transmission, proving submodularity in tree graphs allows application of Sviridenko's [7] result on budgeted maximization of submodular functions. In the multistage-stochastic case, we solve a natural LP with a more complex feasible region than that considered by Ageev & Sviridenko [1], but we are able to extend their pipage-rounding analysis to reduce the number of fractional variables: this requires additional specifications about which pairs of fractional decision variables may be rounded against each other and a careful treatment of the larger number of fractional variables that remain at the end of the pipage stage. All extensions from trees to general graphs employ the probabilistic capacity-preserving mapping of Räcke in the standard way (see [2]): approximate the costs by a distribution over trees, solve a suitably modified instance in each tree, translate solutions back to the original graph, select the best solution. Our techniques also yield similar results for stochastic multistage and probabilistic item-failure extensions of the constrained Maximum Coverage problem.

2 2-Stage Stochastic Graph Protection Problem in Trees

The spread of wild fires can be prevented both through advance fuel treatments and through real-time fire-fighting. Our model captures the tradeoff between using resources in advance vs. waiting until the realization of the ignition point is known but operations are more costly.

The input is a connected tree $T = (V, E)$, a non-negative *value* function $v : V \to \mathbb{Z}$, a non-negative *cost* function $c : E \to \mathbb{Z}$, and a *budget* B. A distribution Π over *source nodes* i is specified. In the first stage Π is known, and it costs c_e to remove edge e from T, in the second phase a realization from Π is specified (say the source is i), and edge e may be removed from T at cost $M^{ie}c_e$. That is: edges purchased in the second stage, once the source is known, have increased cost by a multiplicative *inflation factor* that may depend both on the scenario realized and on the edge.

The total spending on removing edges from T over both phases must be at most B. The objective is to specify a set of edges to buy in the first stage, and then a set of edges to buy in the second stage (depending on the realized source node from Π), such that the expected value not reachable from the realized source node is maximized. We aim to maximize the expected value *protected* from the source. We can contract all edges with costs strictly greater than B since they will not be in any optimal solution.

Special Case 0. Consider the limiting case when all second-stage actions are prohibitively expensive and also Π has support of size 1: this case is the Minimum-Size Bounded-Capacity Cut problem of Hayrapetyan, et al. [4]. They give a PTAS in graphs of bounded tree width and show that this deterministic problem with a single ignition node is NP-hard in trees.

Suppose in this deterministic single-stage case we replace the single ignition point s with a ignition set S. Now the objective is to maximize the expected value protected from every node in S by removing a budget-balanced set of edges.

Theorem 1. *There exists a PTAS in graphs of bounded tree width for the single-stage deterministic Graph Protection Problem (GPP) with a general ignition set.*

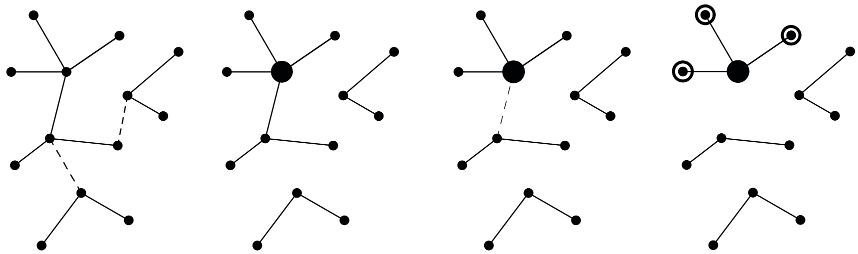

Fig. 1. Leftmost graph: The two dashed edges are removed in stage 1. Second graph: The (bolded) ignition node is realized. Third Graph: After ignition, additional edges can be removed in stage 2. Fourth Graph: Fire spreads through the connected component containing the ignition node: Non-ignition nodes lost to fire are shown circled.

A modified graph with a single source also has bounded tree width, so the existing PTAS can be applied. The PTAS asserted in Theorem 1 for trees can be produced directly by extending the classic dynamic programming framework for Knapsack. (see full paper [6]).

Applying [4] with an enumeration scheme over a polynomial number of divisions of the tree into source-containing components which can each be modified to act as a single-source deterministic problem, we get (details in [6]):

Theorem 2. *There exists a PTAS for the stochastic single-stage GPP in trees provided that the size of the support of the distribution Π and the size of each ignition set given positive weight by Π are bounded by a constant.*

Theorems 1 and 2 can be extended to bicriteria approximations for general graphs as in Engelberg, et. al [2]: the guarantees on value protected (expected value protected) are identical (though δ may be as much as n), and the budget is violated by a $O(\log n)$-factor (applying Räcke's result [5] on cut-capacity approximation). We mention this method briefly at the end of the paper.

Theorem 3. *There exists a bicriteria $(1-(1-\frac{1}{2\delta})^{2\delta}, 2)$-approximation algorithm for the 2-stage stochastic Graph Protection Problem in trees provided that each scenario has a single ignition node (δ denotes the tree diameter).*

In general graphs, for the case of a constant number of scenarios, Theorem 3 can be extended to a $(1-(1-\frac{1}{2n})^{2n}, O(\log n))$-bicriteria approximation (the multistage case requires an application of the Markov inequality to ensure $O(\log n)$-capacity distortion for each scenario under the cut-capacity approximation, details are at the end of the paper).

The following proof of Theorem 3 does not require that the node values are uniform across scenarios, but for notational convenience we will ignore this. This flexibility (and creative use of scenario-dependent edge costs) allows the input form to describe spatial properties of certain types of diffusive processes so that fragmenting the graph has more subtle process-specific implications for value protection than is immediately obvious when considering connectivity (details in [6]).

Roughly, the key ideas of the proof follow: the optimal fractional solution to a natural LP for 2-stage GPP acts as a starting point for a rounding algorithm. The rounding algorithm (carefully) chooses two fractional variables and rounds the LP solution along a vector that maintains their weighted sum (in order to retain feasibility of the budget constraints) while increasing a proxy function that matches the LP objective on integer points and remains boundedly close to the LP objective on fractional points. This is repeated until at most a few fractional variables remain. The effect of some final required roundings can be bounded against the value of an initial partial-enumeration phase. Since the final solution is obtained by a series of increasing steps for the proxy function, it will have high value compared to the original LP solution (for the correct partially-enumerated set). A technical point for the analysis is that a series of such integer solutions must be produced so that the effect of the final required roundings are small. Some simple alterations of this analysis will also yield results for single and k-stage versions as well as for a version in which the first and second stage budgets are specified in the input.

Proof. We formulate the following natural LP: $\max \sum_{(i,v)}(p_i v_v)x_{iv}$ such that $\sum_{e\in P(i,v)} y_e + \sum_{e\in P(i,v)} z_e^i \geq x_{iv}$ for all (i,v) pairs, $\sum_e y_e c_e + \sum_e z_e^i (M^{ie} c_e) \leq B$ for all i, and $x_{iv} \leq 1$ for all (i,v) pairs.

Here, p_i denotes the probability that node i is the ignition point under Π (the scenario where i is the ignition point is *scenario* i). In the associated IP, x_{iv} is 1 if node v is protected in scenario i, and 0 otherwise. Also, y_e is 1 if edge e is bought in the first stage, and 0 otherwise, and z_e^i is 1 if edge e is bought in the second stage for scenario i, and 0 otherwise. Constraints of the first form capture that if node v is protected in scenario i then it must be that some edge on the path from i to v is purchased either in the first stage or in the second stage for scenario i. Constraints of the second form capture that at most B can be spent buying edges in scenario i over the first and second stages combined. Preprocess by setting y_e to 0 if $c_e > B$, and z_e^i to 0 if $M^{ie} c_e > B$: the optimal solution can not use these options. Let δ denote the diameter of the tree.

Notice that in this LP, given a set of y_e and z_e^i, we can automatically determine the best x_{iv}. Following [1] we rewrite the problem as the following nonlinear optimization problem:

$$\max L(x) = \sum_{(i,v)} (p_i v_v) \min\{1, \sum_{e\in P(i,v)} y_e + \sum_{e\in P(i,v)} z_e^i\}$$

$$\text{s.t. } \sum_e y_e c_e + \sum_e z_e^i (M^{ie} c_e) \leq B \quad \forall i, \text{ and } x_{iv} \leq 1 \quad \forall (i,v).$$

Consider the function: $F(x) = \sum_{(i,v)} (p_i v_v) \Big[1 - \Big(\prod_{e\in P(i,v)} (1-y_e)\Big)\Big(\prod_{e\in P(i,v)} (1-z_e^i)\Big)\Big].$

Lemma 1. *$F(x)$ has the following key properties:*

1. *$F(x)$ coincides with $L(x)$ when all the y_e and z_e^i are integral.*
2. *On non-integral (y_e, z_e^i) vectors, $F(x)$ is at least $(1-(1-\frac{1}{2\delta})^{2\delta})L(x)$.*
3. *$F(x)$ is concave in the direction of a vector that changes at most 2 y_e values at a time and changes no $z_{e'}^i$ values. $F(x)$ is concave in the direction of a vector that changes at most 2 z_e^i values for a common i at a time and changes no y_e values, and changes no $z_e^{i'}$ values for $i' \neq i$. Based on the budget constraint coefficients of the changing variables, vectors of this type can be found through appropriate scaling that maintain all budget constraints.*
4. *Let Y, Z denote sets corresponding to the y_e, z_e^i decision variables being set to 1. $F(X)$ defined on subsets of $Y \cup Z$ is a submodular set function.*

Properties 1, 2 and 4 hold just as in [1] since the function $F(x)$ has the same form (though now there is a formal distinction between first and second stage variables). For property 3: the number of terms in F's product which change for any particular (i,v) is at most 2: concavity results as in [1], but unlike in [1], not any set of two fractional decision variables will maintain budget feasibility).

Denote by LP$[I_0, I_1]$ the original LP (post preprocessing) subject to the additional constraints that decision variables in I_1 are set to 1 and decision variables in I_0 are set to 0. We use an auxiliary algorithm $\mathcal{A}$ identical to [1] except for a key additional point. First, $\mathcal{A}$ computes the optimal solution x^{LP} to LP$[I_0, I_1]$ by some known polynomial-time algorithm, then $\mathcal{A}$ transforms this solution into $x^{\mathcal{A}}$ by a series of pipage steps. Each pipage step is as follows. If there exists only a single fractional variable among the y_e, and for every i there is at most a single fractional variable among the z_e^i, stop. Otherwise, select either two fractional y_e or two fractional z_e^i for a common i and consider the vector that maintains all budget constraints as one is increased while the other is decreased: this vector intersects the boundary of the feasibility polytope at two points. At one point the first decision variable has become 0 and the second has become 1, at the other point the second decision variable has become 0 and the first has become 1. Both points are feasible since all budget constraints are maintained, and one has $F(X)$ at least as great as the previous solution due to the concavity of F along the vector. We replace the current solution with this higher-$F(X)$ solution that has a greater number of integral variables.

Each pipage step of $\mathcal{A}$ reduces the number of fractional components of the current vector. Finally $\mathcal{A}$ outputs an *almost-integral* feasible vector $x^{\mathcal{A}}$ which has at most one fractional first-stage variable, and at most one fractional second-stage variable for each scenario i.

As in [1], this rounding procedure gives $F(\mathcal{A}) \geq F(x^{LP})$. Defining $J_1 = \{(i,v) : i \text{ is separated from } v \text{ by } I_1\}$, and from property 2 of the lemma:

$$F(x^{LP}) \geq \sum_{(i,v)\in J_1} p_i v_v + \Big(1-(1-\frac{1}{2\delta})\Big)^{2\delta} \sum_{(i,v)\in J\setminus J_1} (p_i v_v) \min\{1, \sum_{e\in P(i,v)} (y_e)^{LP} + \sum_{e\in P(i,v)} (z_e^i)^{LP}\}$$

Main Algorithm. For each set of at most three y_e, set them to 1, then find the PTAS 2nd stage decision that can be made in each scenario (no additional first

stage edges purchased), and evaluate the objective of each such solution. Take the best such solution and call it q^*.
1. For each $I_1 \subseteq Y$ such that $|I_1| = 4$ and $\sum_{i \in I_1} c_i \leq B$:

- Set $I_0 = \emptyset$.
- Set $t = 0$.
- While $t = 0$: apply $\mathcal{A}$ to LP$[I_0, I_1]$.
  1. If all the $x_i^{\mathcal{A}}$ (decision variables in either stage) are integral, then set t to 1 and set $\hat{x}$ to $x_i^{\mathcal{A}}$.
  2. Else, if $x_i^{\mathcal{A}}$ has no fractional y_e, then round up any fractional z_e^i, set t to 1 and set $\hat{x}$ to $x_i^{\mathcal{A}}$ with the rounded up second stage variables.
  3. Else, if neither of these conditions holds, round down the single fractional y_e and round up all fractional z_e^i, set $\hat{x}$ to $x_i^{\mathcal{A}}$ with the rounded variables. Also, add the index of the y_e that was rounded down to I_0.
  4. If $F(\hat{x}) > F(\bar{x})$, then set $\bar{x}$ to $\hat{x}$. *(Since $\hat{x}$ and $\bar{x}$ are integral, this chooses the highest L-value among all the $\hat{x}$ considered by the algorithm).*

Now we prove that this algorithm meets claim of Theorem 3. First observe that the algorithm spends at most $2B$ for scenario i: pipage rounding maintains budget feasibility for every scenario and the final roundings used to achieve integrality round up at most a single fractional decision variable per scenario. Our preprocessing guarantees that this single round up costs at most B in addition to the cost of the fractional solution returned by $\mathcal{A}$.

Let X^* be the optimal set of decision variables, let Y^* denote the first stage variables in X^*. If $|Y^*| \leq 3$, then step 0. finds a $(1+\epsilon)$ approximation to OPT. So, we address the case when $|Y^*| \geq 4$. W.l.o.g. we can assume that the set of decision variables is ordered such that $Y^* = \{1, ..., |Y^*|\}$ and for each $i \in Y^*$, among the elements $\{i, ..., |Y^*|\}$ the element i protects the maximum total weight of (i, v) pairs which are not already protected by the set $\{1, ..., i-1\}$.

For the iteration in which $I_1 = \{1, 2, 3, 4\}$, let q denote the number of runs of the while loop. Since each run of the while loop either terminates the iteration or sets a first stage variable to 0, q is at most $n - 4$. During the iteration the algorithm finds a series of q feasible solutions to the LP. Let I_0^j denote I_0 in the jth run of the while loop. The jth feasible solution $\hat{X}_j$ has $\hat{X}_j \cap I_0^j = \emptyset$ (from the form of the algorithm). Index the elements of I_0^q in the order that the algorithm adds them to I_0, that is, $I_0^j = \{i_1, ..., i_j\}$ where i_l is the index of the lth first stage variable added to I_0 for this iteration.

Assume first that $I_0^q \cap Y^* = \emptyset$. That is, when the iteration terminates, no first stage variables used by OPT have been forced to 0: OPT is a feasible solution for LP$[I_1, I_0^q]$. Since this is the last run of the while loop, it must have ended in an *if* statement of one of the first 2 types. In the first case: all the $x_i^{\mathcal{A}}$ (decision variables in either stage) are integral and $x_i^{\mathcal{A}}$ is the outcome of pipage rounding of the fractional optimal of LP$[I_1, I_0^q]$. In particular: since $\hat{x}$ is integral, $L(\hat{x}) = F(\hat{x}) = F(x_i^{\mathcal{A}}) \geq F(x^{LP})$. For the second case, rounding up the second-stage variables only increases the value of F, and after the rounding we have an integral solution, so $L(\hat{x}) = F(\hat{x}) \geq F(x_i^{\mathcal{A}}) \geq F(x^{LP})$. Either way, the following inequality derived from property 2 and the fact that OPT is feasible for LP$[I_1, I_0^q]$ now gives that $\hat{x}$ is a budget-balanced $(1-(1-\frac{1}{2\delta})^{2\delta})$-approximation:

$$F(x^{LP}) \geq \sum_{(i,v)\in J_1} p_i v_v + \left(1-(1-\frac{1}{2\delta})\right)^{2\delta} \sum_{(i,v)\in J\setminus J_1} (p_i v_v) \min\{1, \sum_{e\in P(i,v)} (y_e)^{LP} + \sum_{e\in P(i,v)} (z_e^i)^{LP}\}$$
$$\geq \left(1-(1-\frac{1}{2\delta})\right)^{2\delta} OPT.$$

Now, assume that $I_0^q \cap Y^* \neq \emptyset$. Let I_0^{s+1} be the first I_0 in the series $I_0^1, ..., I_0^q$ that has nonempty intersection with Y^*: the sth run of the while loop is the first run of the while loop for this iteration in which the algorithm adds a first stage variable from Y^* to I_0 (call that variable i_s). The algorithm adds i_s to I_0 after considering a solution $\hat{x}$ in which i_s was the single fractional first stage variable was rounded down (this is the third type of *if* statement in the while loop). We claim that the $\hat{x}$ that resulted when i_s was rounded down (and fractional second stage variables were rounded up) was a $(1-(1-1/2\delta)^{2\delta})$-approximation. Proving this claim will be establish Theorem 3.

As in [1], $F(X)$ defined on subsets of $Y \cup Z$ is a submodular set function. Thus, we have the *diminishing-returns* property: for any subsets R and G of $Y \cup Z$ and any element $i \in Y \cup Z$, we get $F(R \cup i) - F(R) \geq F(R \cup G \cup i) - F(R \cup G)$. Now, letting h denote a member of Y^* which is not in $\{1, 2, 3, 4\}$, and letting H denote any superset of $\{1, 2, 3, 4\}$:

$$1/4F(I_1) = 1/4F(1,2,3,4)$$
$$= 1/4[F(\{1,2,3,4\}) - F(\{1,2,3\}) + F(\{1,2,3\}) - F(\{1,2\}) + F(\{1,2\}) - F(\{1\}) + F(\{1\}) - F(\emptyset)]$$
$$\geq 1/4[F(\{1,2,3,h\}) - F(\{1,2,3\}) + F(\{1,2,h\}) - F(\{1,2\}) + F(\{1,h\}) - F(\{1\}) + F(\{h\}) - F(\emptyset)]$$
$$\geq F(H \cup \{h\}) - F(H).$$

The first equality results from a collapsing sum where we remove the final $+F(\emptyset)$ since it is 0 (since the tree is connected and every scenario has a source). By the labeling of the decision variables in Y^*: since h is not in $\{1, 2, 3, 4\}$, the additional marginal value h protects beyond what is protected by any prefix of $\{1, 2, 3, 4\}$ is at most the additional value that the index which does follow the prefix protects. Finally, we apply the diminishing-returns property 4 times to get the final inequality.

Also, as in [1], rounding up a fractional solution produced by $\mathcal{A}$ only increases the value of F. Let $x^{\mathcal{A}}$ denote the unrounded solution returned by $\mathcal{A}$. Let $I(x^{\mathcal{A}})$ be the integral positive elements of $x^{\mathcal{A}}$, let $\{j_1, ..., j_i\}$ denote the set of fractional second stage variables in $x^{\mathcal{A}}$, and i_s denote the fractional first stage variable in $x^{\mathcal{A}}$ from Y^*. Then $\hat{x}$ is $I(x^{\mathcal{A}}) \cup \{j_1, ..., j_i\}$, so we can use the integrality of $\hat{x}$ to bound its LP value as follows:

$$L(\hat{x}) = L(I(x^{\mathcal{A}}) \cup \{j_1, ..., j_i\}) = F(I(x^{\mathcal{A}}) \cup \{j_1, ..., j_i\})$$

Adding and subtracting a common quantity:

$$= F(I(x^{\mathcal{A}}) \cup \{j_1, ..., j_i\} \cup \{i_s\}) - \left(\underbrace{F(I(x^{\mathcal{A}}) \cup \{j_1, ..., j_i\} \cup \{i_s\}) - F(I(x^{\mathcal{A}}) \cup \{j_1, ..., j_i\})}\right)$$

Applying our bound to bracketed quantity since $I(x^{\mathcal{A}})$ contains $\{1,2,3,4\}$ and $i_s \in Y^*$:

$$\geq F(I(x^{\mathcal{A}}) \cup \{j_1, ..., j_i\} \cup \{i_s\}) - 1/4F(I_1) \geq F(x^{\mathcal{A}}) - 1/4F(I_1)$$

The second inequality holds because F increases when its argument is rounded up, and $I(x^{\mathcal{A}}) \cup \{j_1, ..., j_i\} \cup \{i_s\}$ is just $x^{\mathcal{A}}$ rounded up. Now write out $F(x^{\mathcal{A}})$:

$$= \sum_{(i,v)\in J_1} p_i v_v + \sum_{(i,v)\in J\setminus J_1} (p_i v_v)\Big[1 - \Big(\prod_{e\in P(i,v)} (1-(y_e)^{\mathcal{A}})\Big)\Big(\prod_{e\in P(i,v)} (1-(z_e^i)^{\mathcal{A}})\Big)\Big] - 1/4F(I_1)$$
$$= 3/4 \sum_{(i,v)\in J_1} p_i v_v + \sum_{(i,v)\in J\setminus J_1} (p_i v_v)\Big[1 - \Big(\prod_{e\in P(i,v)} (1-(y_e)^{\mathcal{A}})\Big)\Big(\prod_{e\in P(i,v)} (1-(z_e^i)^{\mathcal{A}})\Big)\Big]$$

Pipage rounding produces $x^{\mathcal{A}}$ from x^{LP} while increasing F:

$$\geq 3/4 \sum_{(i,v)\in J_1} p_i v_v + \sum_{(i,v)\in J\setminus J_1} (p_i v_v)\Big[1 - \Big(\prod_{e\in P(i,v)} (1-(y_e)^{LP})\Big)\Big(\prod_{e\in P(i,v)} (1-(z_e^i)^{LP})\Big)\Big]$$

Apply the well-known inequality which holds for all fractional solutions:

$$\geq 3/4 \sum_{(i,v)\in J_1} p_i v_v + (1-(1-1/2\delta)^{2\delta}) \sum_{(i,v)\in J\setminus J_1} (p_i v_v) \min\{1, \sum_{e\in P(i,v)} (y_e)^{LP} + \sum_{e\in P(i,v)} (z_e^i)^{LP}\}$$

Notice that $3/4 \geq (1-(1-1/2\delta)^{2\delta})$. Also, x^{LP} is the optimal solution for LP$[I_1, I_0^s]$ and X^* is feasible for LP$[I_1, I_0^s]$. Thus, the last quantity is bounded below by $(1-(1-1/2\delta)^{2\delta})L(X^*) = (1-(1-1/2\delta)^{2\delta})OPT$.

Suppose that the division of the budget between first and second stages is specified in the input as (B_1, B_2). Adding the additional constraints $\sum_e y_e c_e \leq B_1$ and $\sum_e z_e^i(M^{ie}c_e) \leq B_2$ for all i to the LP alters our analysis only slightly: preprocess to eliminate decision variables that are too expensive to fully buy in their corresponding stages, the algorithm now enumerates over four-member sets of first-stage decision variables, at the conclusion of the pipage phase the remaining fractional first-stage variable is rounded down (so B_1 is respected) and at most one second-stage variable per scenario is rounded up (B_2 is overspent by at most a factor of 2), first stage variables which are rounded down are excluded one by one in the iterations of the while loop. Thus, we get:

Theorem 4. *Given a specific first-stage budget B_1 and second-stage budget B_2, there exists a $(1-(1-\frac{1}{2\delta})^{2\delta})$-approximation algorithm for the 2-stage stochastic GPP in trees that respects B_1 and violates B_2 by a factor of at most 2 (each ignition set has size 1, δ denotes the diameter of the tree).*

Stochastic Single-Stage and k-Stage Results. In the limiting single-stage stochastic case (where second-stage action is prohibitively expensive) there is only a single budget constraint: the proof of Theorem 3 can be simplified so that it directly follows [1] to get:

Theorem 5. *There exists a $(1-(1-\frac{1}{\delta})^{\delta})$-approximation algorithm for the single-stage stochastic GPP in trees provided that each ignition set has size 1 (δ denotes the diameter of the tree).*

For the 2-stage stochastic GPP in trees with single ignition node, consider the algorithm that chooses the better performance between spending all of B in stage 1 vs. spending all of B in stage 2: apply Theorem 5 for stage 1 and the PTAS for deterministic single-source GPP for stage 2 assuming that the optimal solution earns $\alpha(\text{OPT})$ in the first stage, and minimize over $\alpha \in (0,1)$ to get a worst case guarantee of $(0.387, 1)$. For a constant number of scenarios, use Theorem 2 in the place of Theorem 5 to get a $(.5(1-\epsilon), 1)$- approximation.

The k-stage stochastic graph protection problem in trees (for constant k) has k stages in which information is revealed and decisions about edge removal are made (rather than one or two stages). This information can be considered as updates that arrive at k specific times which condition the distribution on where the ignition will occur (by specifying that the ignition will occur among some particular subset of the nodes). For each stage the input includes a partition of the node set, and the partition for stage i refines the partition for stage $i-1$. In each stage the planner has the option to remove additional edges from the graph at some (stage, partition piece)-specific cost. A solution specifies which edges will be removed for each partition piece realization at each stage. The total cost incurred for each realized sequence of k partition pieces should be B.

Theorem 6. *For a restricted class of information revelation hierarchies, there exists a bicriteria* $(1-(1-\frac{1}{k\delta})^{k\delta}),\ 2+\epsilon)$*-approximation algorithm for the k-stage stochastic GPP in trees provided that each ignition set has size 1 (k is a constant,* δ *denotes the diameter of the tree).*

Theorem 6 requires that the number of partition pieces added over all stages excluding the last stage (in which any of n points may be realized) is bounded by a constant: guessing the optimal division of the budget to ϵ/k-precision for each possible information realization takes polynomial time. As in the (B_1, B_2) case: impose additional constraints based on the guess of optimal budget division, reject too-expensive decision variables, pipage round (now roundings take place between pairs of fractional variables that correspond to a common partition piece within a stage). Last, round up all fractional variables (see [6] for details).

If there is a specified budget for each of the k stages, then the guessing (enumeration) may be dropped: with no requirements on the information revelation hierarchy the same analysis gives a $(1-(1-\frac{1}{k\delta})^{k\delta})$ value-protection guarantee which violates each stage's budget by a factor of at most 2.

Reductions, Results for Stochastic Multistage MCKP. A looser $(1-\frac{1}{e})$ guarantee which matches Theorem 5 asymptotically may be obtained by reducing single-stage stochastic GPP in trees to the weighted Budgeted Separating Minimum Cut Problem in trees for which the analysis of Engelberg, et. al [2] applies: submodularity of the objective allows application of the result of Sviridenko [7]). The tighter result in Theorem 5 can alternately be proved by a more subtle reduction to MCKP addressed in [1] (reducing wBSMC in trees to MCKP gives the tighter result for wBSMC as well). Full Reductions in [6].

Maximum Coverage with a Knapsack Constraint (MCKP): Given a family $F = \{S_j : j \in J\}$ of subsets of a set $I = \{1, 2, ..., n\}$ with associated nonnegative weights w_j and costs c_j of the elements, and positive integer B, find a subset $X \subseteq I$ with $\sum_{j \in X} c_j \leq B$ so as to maximize the total weight of the sets in F having nonnegative intersections with X.

- *Stochastic MCKP:* There is also a distribution Π: each scenario specifies how much value will be received for covering the subset S_j for each j. The objective is to maximize the expected weight of subsets covered.
- *Multistage MCKP:* Elements may be purchased in different stages at a cost that is stage-, scenario-, and element-dependent (costs are specified in the input). Stochastic multistage versions of wBSMC in trees reduce to these MCKP problems.

The features of the LP we analyzed (objective function and budget constraints) also hold for the natural LPs for these problems: the analysis proving theorems 3, 4, 5, and 6 can be extended with identical guarantees to the corresponding multistage stochastic MCKP generalizations.

3 1-Stage Extension to Probabilistic Edge Transmission

In ecological fact, fuel-treated areas are not 100% burn resistant (e.g. they may burn if extreme weather arises). Also, different types of treatments (with different costs) may reduce the probability of fire passing between adjacent parcels by different amounts. These considerations motivate a version of GPP in which the input specifies a more complicated relationship between spending on each edge and the resulting transmission probability across that edge. Previously we had two options: pay 100% of the edge cost to get probability of transmission 0, or pay 0% of the edge cost to get probability of transmission 1.

To single-stage stochastic GPP where each ignition set has size 1, we add the feature that each edge has (as part of the input) a specified monotonically-decreasing step function that gives the probability of transmission across that edge as a function of the spending level (the spending level may range from 0% to 100% of the edge cost, the events of transmissions across edges are assumed to be independent). We give an approximation result assuming that the running time of the algorithm is allowed to depend polynomially on the number of steps in each step function. The objective remains to maximize the expected value protected from the ignition point, only now this expectation is over realization of both the scenario and the individual edge-transmission events that arise.

The analogous notion for MCKP is of *probabilistic element failure*: for each element there is a step function that represents the probability that the element will *fail to cover the subsets which contain it* (generalizing that an element e fails to cover subsets which contain it with probability 1 if we do nothing, and with probability 0 if we pay c_e). The objective is to maximize the expected weight of subsets covered, where this expectation is over both element and scenario realization. The generalization of wBSMC in trees to a case with probabilistic edge occurrence reduces to MCKP with *probabilistic element failure*.

Theorem 7. *There exists a* $(1-\frac{1}{e})$*-approximation algorithm for the single-stage stochastic GPP in trees with probabilistic edge transmission (provided that each ignition set has size 1). For MCKP with probabilistic element failure: there exists a* $(1-\frac{1}{e})$*-approximation algorithm.*

Proof (GPP). Each (spending level, edge) pair is an element the solution can buy with cost corresponding to the spending level times the edge cost (we only have elements corresponding to critical spending levels at which the transmission probability instantaneously drops). Let X denote the set of such elements. The expected value protected is a set function over these elements. Denote this function by E. We wish to maximize this set function by buying elements subject to a knapsack constraint: if we show that this set function is submodular, [7] will immediately yield a $(1-\frac{1}{e})$-approximation that is budget-balanced (provided that we can compute in polynomial time the element which gives largest improvement). To prove submodularity we will establish the law of diminishing returns: for an arbitrary (spending level, edge) pair denoted by a, if $A \subseteq B \subseteq X$, then $E(A \cup a) - E(A) \geq E(B \cup a) - E(B)$.

Let the edge of the (spending level, edge) pair a be denoted by e. According to the step function for e, buying a results in some probability of transmission α_i. Before a is added, A contains some set of elements which affect the transmission probability on e, and B contains a superset of these elements. Thus the probability of transmission on e is (weakly) larger for the set A than for the set B. In both cases, when a is added to a set, the new probability of transmission on e is the minimum of α_i and the current probability of transmission on e. The gap is larger for A than for B. Let $\wp_e(\cdot)$ denote the probability of transmission on e as a function of the set of elements: $\wp_e(A) \geq \wp_e(B) \Rightarrow \wp_e(A) - \wp_e(A \cup a) \geq \wp_e(B) - \wp_e(B \cup a)$.

Next, focus on a particular (ignition point, node) pair (i, v). If the path from i to v does not contain e, then adding e does not change the (i, v)th term in the expression for expected value protected. If the path from i to v does contain e, for each non-e edge on this i to v path, the probability of transmission under A is at least the probability of transmission under B. Let $P(Q)$ denote the probability that every edge on the i to v path (excluding e) transmits under Q:

$$
\begin{aligned}
&P(A) \geq P(B) \Rightarrow P(A)(\wp_e(A) - \wp_e(A \cup a)) \geq P(B)(\wp_e(B) - \wp_e(B \cup a)) \Rightarrow \\
&\quad\Rightarrow P(A)\wp_e(A) - P(A)\wp_e(A \cup a) \geq P(B)\wp_e(B) - P(B)\wp_e(B \cup a). \\
&\quad\Rightarrow (1 - P(A)\wp_e(A \cup a)) - (1 - P(A)\wp_e(A)) \geq (1 - P(B)\wp_e(B \cup a)) - (1 - P(B)\wp_e(B)). \\
&\quad\Rightarrow E(A \cup a) - E(A) \geq E(B \cup a) - E(B).
\end{aligned}
$$

The third line compares the changes in probability that v is protected from i which result when a is added to A and when a is added to B. The final inequality follows from summing change in expected valued protected over (ignition point, node) pairs (including pairs for which the addition of e caused no change). This establishes submodularity. Computing the change in E resulting from the addition of a single element simply requires computing the product along the (ignition point, node) path twice for each (i, v) pair. This takes polynomial time for each of polynomially-many elements.□.(Similar MCKP analysis in [6]).

Extensions to General Graphs

Single-Stage. Theorems 1, 2, 5 can be extended to bicriteria approximations for general graphs: the guarantees on value protected (expected value protected) are identical, and the budget is violated by a $O(\log n)$-factor. As in Engelberg, et. al [2] we apply the result of Räcke [5] on cut-capacity approximation: approximate the costs graph by a distribution over tree graphs (whose maximum diameter is n), solve a suitably modified instance in each tree, translate solutions back to the original graph, select the best solution.

Multi-stage. If the number of scenarios is bounded by a constant, then Theorems 3, 4, and 6 can be extended to general graphs: the guarantees on expected value protected are identical, but the budget(s) is violated by a $O(\log n)$-factor. To apply the result of Räcke [5] to the multistage case we need that some tree produced by the cut-capacity approximation has $O(\log n)$-distortion *for the optimal solution in every scenario* (not just for a single set of edges purchased in the first stage). Details in [6].

References

1. Ageev, A.A., Sviridenko, M.: Pipage rounding: A new method of constructing algorithms with proven performance guarantee. J. Comb. Optim. 8(3), 307–328 (2004)
2. Engelberg, R., Könemann, J., Leonardi, S., Naor, J(S.): Cut Problems in Graphs with a Budget Constraint. In: Correa, J.R., Hevia, A., Kiwi, M. (eds.) LATIN 2006. LNCS, vol. 3887, pp. 435–446. Springer, Heidelberg (2006)
3. Finney, M.A.: A computational method for optimising fuel treatment locations. International Journal of Wildland Fire 16, 702–711 (2007)
4. Hayrapetyan, A., Kempe, D., Pál, M., Svitkina, Z.: Unbalanced Graph Cuts. In: Brodal, G.S., Leonardi, S. (eds.) ESA 2005. LNCS, vol. 3669, pp. 191–202. Springer, Heidelberg (2005)
5. Räcke, H.: Optimal hierarchical decompositions for congestion minimization in networks. In: 40th STOC, pp. 255–264 (2008)
6. Shmoys, D., Spencer, G.: Full paper: Approximation algorithms for fragmenting a graph against a stochastically-located threat, `http://people.orie.cornell.edu/gms39/images/documents/gsfragment.pdf`
7. Sviridenko, M.: A note on maximizing a submodular set function subject to a knapsack constraint. Operations Research Letters 32, 41–43 (2004)
8. Wei, Y., Rideout, D., Kirsch, A.: An opt. model for locating fuel treatments across a landscape to reduce expected fire losses. Can. J. of Forest Res. 38, 868–877 (2008)