Computational Aspects of Market-Clearing Mechanisms in Deregulated Power Industry*

DOUG COOK† VANCE FABER‡ MADHAV MARATHE§ ARAVIND SRINIVASAN¶
YORAM J. SUSSMANN‖

Abstract

We consider combinatorial problems associated with contract satisfaction and market clearing arising in deregulated electrical power industry. A prototypical problem in this context can be stated as follows: Given a network G and a (multi-)set of pairs of vertices in it denoting bilateral contracts, find the maximum number of simultaneously satisfiable contracts. The extension of the above problem to Poolco-type contract satisfaction problems is considered. Such problems also arise in real-time Internet services (e.g., telephone, fax, video).

We show that these problems come in a few variants, some efficiently solvable and many *NP*-hard; we also present approximation algorithms for many of the *NP*-hard variants presented. Some of our approximation algorithms benefit from certain improved tail estimates that we derive; the latter also yield improved approximations for a family of packing integer programs.

Key words and phrases: Combinatorial optimization, production and transmission of power, regulation of electric power industry, approximation algorithms, splittable flow, routing algorithms.

1 Introduction

The U.S. electric utility industry is in the early stages of major structural changes driven by the move to deregulate the industry [7, 17, 19, 20, 47]. A major consequence of deregulation is that consumers as well as producers will eventually be able to negotiate prices to buy and sell electricity. See the comprehensive discussions in [45, 46, 19, 47] for more details on this topic. Before formally defining the problems, we view the setting informally for now as a collection of request pairs (*contracts*) in a flow network wherein the flow for any pair can be *split into multiple paths*. In practice, deregulation is complicated by the fact that all power companies will have to share the same power network in the short term, with the network's capacity being

*The work of all authors except for Aravind Srinivasan has been supported by EPRI and Department of Energy under Contract W-7405-ENG-36. A preliminary version of this paper entitled *Low Bandwidth Routing and Electrical Power Networks* appeared in *Proc. 25th International Colloquium on Automata, Languages and Programming (ICALP)*, Aalborg, Denmark, LNCS 1443, Springer Verlag, pp. 604-615, 1998.

†Work done while visiting Los Alamos National Laboratory, Los Alamos NM 87545. Department of Engineering, Colorado School of Mines, Golden CO 80401.

‡Lizardtech Inc. Work done while at Los Alamos National Laboratory.

§Corresponding author. Current address: P.O. Box 1663, MS B265, Los Alamos National Laboratory, Los Alamos NM 87545. Email: `marathe@lanl.gov`.

¶Bell Laboratories, Lucent Technologies, 600 Mountain Ave., Murray Hill, NJ 07974-0636. This author's work was done while at the School of Computing, National University of Singapore, Singapore 119260, and supported in part by National University of Singapore Academic Research Fund Grant RP970607. Email: `srin@research.bell-labs.com`.

‖Part of this work was done while visiting Los Alamos National Laboratory, Los Alamos NM 87545. Supported in part by NSF CAREER Award CCR-9501355. Email: `yorams@atl.fundtech.com`.

just about sufficient to meet current demands. Under deregulation, most U.S. states are planning to set up an independent system operator (ISO), a governing body to arbitrate the use of the network. The basic questions facing the ISOs will be how to decide which contracts to deny/accept (due to capacity constraints), and who is to bear the costs involved in such denials. Such market-clearing mechanisms will play a crucial role if the deregulated power market were to eventually become *economically efficient*: i.e., prices are fair and no individual or a group of consumers/suppliers can execute *market power*. Several criteria/policies have been proposed and/or are being legislated by the states as possible guidelines for the ISO for finding the "best" set of feasible contracts [45]. These include:

Last in, First Out: In this plan, all contracts are registered with the ISO in the order they are received. Any contract that results in an infeasibility of the resulting solution is denied. As argued comprehensively by Wildberger [45], such a policy typically favors forward contracts; although in some cases it favors producers who opted for spot prices.

Minimum Flow denied: Under this proposal, the ISO considers all contracts simultaneously and selects the combination that denies the *least* amount of proposed power flow.

Minimum Total Cost: Similar to **Minimum Flow denied** except that the highest priced contracts will be eliminated or adjusted down first. This plan attempts to benefit the consumers by minimizing their cost.

Maximum Total Cost: Similar to **Minimum Flow denied** but retain the contracts with maximum total cost. This is being advocated by legislatures in view of the anticipated tax revenues. It is likely to benefit power producers and also owners of the transmission system but could be disadvantageous to consumers.

The above discussion and our experience in combinatorial optimization suggests that the following important additional parameters will come into play as a result of deregulation: (i) the underlying network, (ii) its capacity and topology and (iii) the spatial locations of the bilateral/Poolco contracts on this network. (The Poolco model is defined in Section 6.1.)

This paper investigates the *computational complexity* of executing some of the above-stated policies by the ISO in the event the existing set of contracts result in an infeasible situation for the transmission network. The important goals of this work include: (i) providing a quantitative justification for selecting one policy over another (solely on the basis of computational complexity), (ii) demonstrating that the above parameters crucially affect the way power is routed in the network and that these constraints make the problems at hand much harder than the traditional problems of optimal scheduling, and (iii) whenever possible, providing approximation algorithms with worst-case guarantees for implementing the policies. We elaborate on these results further in Section 2.1. Much research has been conducted in electrical engineering, on the general area considered here [35, 51, 11, 4, 5, 36, 49, 50]. However, these works do not address the issue of contracts considered here. Most of the issues addressed in the past were concerned with finding solutions to the Unit Commitment and Economic Dispatch problems (see [50, 25] and the references therein). The main parameters considered by the authors were: (i) planning over a given time period, (ii) setup costs involved in bringing the power units to life and then shutting them off, etc. Researchers have also used techniques from Artificial Intelligence to solve Unit Commitment and related problems (see [22] and references therein).

Low-bandwidth routing in communication networks. The above setting is also applicable in the context of packet routing in telecommunication networks. Several telecommunications companies are devising protocols that subdivide audio and video signals into smaller packets and reassemble them at the destination, for real-time Internet services (e.g. phone, fax, etc.) [18, 34]. These problems has spurred much attention on the classical *NP*-hard *maximum edge-disjoint-paths problem* (MDP): given a graph G and a (multi-)set of pairs of vertices in it, connect as many of the given pairs as possible using edge-disjoint paths in G [30]. The emergence of high-bandwidth networks supporting heterogeneous applications has also led to a generalization

of the MDP to the *unsplittable flow problem* (UFP): each network link has a capacity, each request pair has a demand, and to satisfy a request, *all* of its demand must be routed through a single path [31, 33, 43]. The usual and natural assumption made here is that no single demand exceeds any capacity. What about the generalization of the MDP in the other direction to the *low bandwidth* case, where large demands must sometimes be serviced by a fixed network? Since some demands may exceed the link capacities, the flows for some request pairs will have to be split into different flow-paths. The results obtained reveal some striking differences between the problems considered here and the MDP/UFP.

Organization. In Section 2 we present the combinatorial optimization problems considered in this paper. The formal problems are an abstraction of the problems that would arise as a result of implementing several of the above-stated policies. Section 2 also summarizes our results and discusses related work. Section 3 contains illustrative examples that provide insights into the problem structure as well as potential subtleties that might arise in a deregulated environment. Section 4 outlines the computational intractability results. Section 5 contains useful new probabilistic tools that might be of independent interest. In Section 6 we propose the new market-clearing mechanism and in Section 7 describe our approximation algorithm for this problem. This yields as a direct corollary an approximation algorithm for the single source version of the (0/1-VERSION, MAX-#CONTRACTS) considered in Section 2. Section 8 discusses the extension of our results to other variants and also outlines how the results here can be used to improve upon the performance of approximation algorithms for certain packing problems. Finally, Section 9 contains concluding remarks and directions for future research.

2 Problem Definitions and Results

The variants of flow problems related to power transmission studied here are intuitively harder than some traditional multi-commodity flow problems, since although the flow out of a given source must equal the flow into the corresponding sink, we *cannot distinguish between* the flow "commodities" (power produced by different generators). See below and Section 3 for more on this. As a result, standard solution techniques used to solve single/multi-commodity flow problems are not directly applicable to the problems considered here. We will use a new rounding technique that gives good approximation bounds. We shall use power terminology throughout, but all results will hold for the message/voice-data routing domains discussed above. In particular, it is easy to modify our algorithms for such multi-commodity cases where the flows for the different commodities are *distinguishable.*

The basic setting is as follows. We are given an undirected network (the power network) $G = (V, E)$ with capacities c_e for each edge e and a set $C = \{(s_1, t_1), (s_2, t_2), \ldots, (s_k, t_k)\}$ of source-sink node pairs. Each pair (s_i, t_i) has: (i) an integral *demand* reflecting the amount of power that s_i agrees to supply to t_i and (ii) a negotiated *cost* of sending unit commodity from s_i to t_i. As is traditional in the power literature, we will refer to the source-sink pairs along with the associated demands as *a set of contracts*. In general, a source or sink may have multiple associated contracts. We will find the following notation convenient to describe the problems. The contracts are defined by a relation $R \subseteq (V \times V \times \Re \times \Re)$ so that tuple $(v, w, \alpha, \beta) \in R$ denotes a contract between source v and sink w for α units of commodity at a cost of β per unit of the commodity. For $A = (v, w, \alpha, \beta) \in R$ we denote $source(A) = v$, $sink(A) = w$, $flow(A) = \alpha$ and $cost(A) = \beta$. We construct a digraph $H = (V \cup S \cup T \cup \{s, t\}, E')$ with source s, sink node t, capacities $u : E' \to \Re$ and costs $c' : E' \to \Re$ as follows. For all $A \in R$, define new vertices v_A and w_A. Let $S = \{v_A \mid A \in R\}$ and $T = \{w_A \mid A \in R\}$. Each edge $\{x, y\}$ from G is present in H as the two arcs (x, y) and (y, x) that have the same capacity as $\{x, y\}$ has in G, and with cost 0. In addition, for all $A = (v, w, \alpha, \beta) \in R$, we introduce: (i) arcs (v_A, v) and (w, w_A) with infinite capacity and zero cost; (ii) arc (s, v_A) with capacity $flow(A) = \alpha$

and cost 0; and (iii) arc (w_A, t) with capacity $flow(A) = \alpha$ and cost equaling $cost(A)$. A *flow* is simply an assignment of values to the edges in a graph, where the value of an edge is the amount of flow traveling on that edge. The value of the flow is defined as the amount of flow coming out of s (or equivalently the amount of flow coming in to t). A generic *feasible flow* $f = (f_{x,y} \geq 0 : (x, y) \in E')$ in H is any non-negative flow that: (a) respects the arc capacities, (b) has s as the only source of flow and t as the only sink. Note that for a given $A \in R$, in general it is not necessary that $f_{s,v_A} = f_{w_A,t}$. For a given contract $A \in R$, A is said to be *satisfied* if the feasible flow f in H has the additional property that for $A = (v, w, \alpha, \beta)$, $f_{s,v_A} = f_{w_A,t} = \alpha$; i.e., the contractual obligation of α units of commodity is shipped out of v and the same amount is received at w. A contract set R is *feasible* (or *satisfied*) if there exists a feasible flow f in the digraph H that satisfies every contract $A \in R$. Let $B = supply(s) = demand(t) = \sum_{A \in R} flow(A)$.

Definition 1 *Given a graph $G(V, E)$ and a contract set R, the* R-MAX-FLOW *problem is to determine if R is feasible.*

In practice, it is typically the case that R does not form a feasible set. As a result we have two possible alternative methods of relaxing the constraints: (i) relax the notion of feasibility of a contract and (ii) find a subset of contracts that are feasible. Combining these two alternatives, we define the following types of "relaxed feasible" subsets of R.

Definition 2 *Let $G(V, E)$ be a power network, R be a set of contracts, H be the associated digraph, and f be a feasible flow in H.*

1. *A contract set $R' \subseteq R$ is a* 0/1-contract satisfaction *feasible set if, $\forall A = (v, w, \alpha, \beta) \in R'$, $f_{s,v_A} = f_{w_A,t} = \alpha$.*
2. *A contract set $R' \subseteq R$ is an* I-contract satisfaction *feasible set if, $\forall A = (v, w, \alpha, \beta) \in R'$, $f(A) := f_{s,v_A} = f_{w_A,t} \in \{0, 1, \ldots, \alpha\}$; i.e., we must send an integral amount of flow $f(A)$ from v to w.*
3. *A contract set $R' \subseteq R$ is an* R-contract satisfaction *feasible set if, $\forall A = (v, w, \alpha, \beta) \in R'$, $f(A) := f_{s,v_A} = f_{w_A,t} \in [0, \alpha]$; i.e., we are allowed to send any rational amount of flow $f(A)$ from v to w.*

Remarks: Note that case (3) of Definition 2 is the least restrictive; the only requirement we have is that the source-destination pairs send and receive equal amounts of flows. Also, all our definitions include at the very minimum a balancing constraint for satisfied (feasible) contracts. For the remaining contracts, the above definitions do not impose any requirement as long as we have a feasible flow f. Note also that given a flow f in H, it is easy to recover the "relaxed feasible" set R' according to any of the above given criteria in polynomial time.

Definition 3 *Given a graph $G(V, E)$ and a contract set R, the (*R-VERSION, MAX-FEASIBLE FLOW*) (resp. (*0/1-VERSION, MAX-FEASIBLE FLOW*), (*I-VERSION, MAX-FEASIBLE FLOW*)) problem is to find a feasible flow f in H such that $\sum_{A \in R'} f(A)$ is maximized where R' forms an* R-contract satisfaction *(resp.* 0/1-contract satisfaction, I-contract satisfaction*) feasible set of contracts.*

Observe that (R-VERSION, MAX-FEASIBLE FLOW) and its weighted version can be written as linear programs and hence solved in polynomial time. In contrast, we shall see that they become intractable when some further restrictions are placed on the structure of feasible solutions. The next class of problems aim at maximizing the number of satisfied contracts.

Definition 4 *Given a graph $G(V, E)$ and a contract set R, the (*0/1-VERSION, MAX-#CONTRACTS*) problem is to find a feasible flow f in H such that $|R'|$ is maximized where R' forms a* 0/1-contract satisfaction *feasible set of contracts.*

The integer and rational variants of the problem [(I-VERSION, MAX-#CONTRACTS) and (R-VERSION, MAX-#CONTRACTS)] are not as interesting since we can satisfy small demands and claim that a contract is satisfied. We can impose an additional condition that for a contract to be satisfied, at least a certain fraction of the demand must be met. We shall study these nontrivial variants later. A natural restriction of the above problems is to have all the sources s_i to share a common node, i.e., $\forall i, s_i = s$ for some $s \in V$. Given a problem Π above we denote the restriction of Π to the case when all the sources share a vertex by SINGLE-SOURCE-Π. The weighted version for each of the problems Π denoted by WT$-\Pi$ is the same as Π, except that each contract $A \in R$ has a desired weight (profit) denoting its importance.

Note that each problem above directly corresponds to a possible policy that might be used to process the contracts. See [3, 10, 9, 25, 50] for basic definitions in computational complexity, and for concepts related to the generation, operation and control of electric power.

2.1 Summary of Results

For the first time in the literature, we study the complexity and approximability of several contract satisfaction problems. Where possible, we state the hardness results for the most restricted versions and approximation results for the most general versions of the problems. Given the flow network $G = (V, E)$, we let $n = |V|$ and $m = |E|$. Recall that an approximation algorithm for an optimization problem Π provides a **performance guarantee** of ρ if for every instance I of Π, the value p returned by the approximation algorithm is within a factor ρ of the optimal value OPT for I: i.e., $p \leq \rho \cdot OPT$ if Π is a minimization problem, and $p \geq OPT/\rho$ if Π is a maximization problem.

Our first main result is for the *single source* version of (0/1-VERSION, MAX-#CONTRACTS). We show that unless *NP* $\subseteq ZPP$, no polynomial-time algorithm can guarantee an approximation factor of $m^{\frac{1}{2}-\epsilon}$ for *any* fixed $\epsilon > 0$, even if all capacities are 1 and all demands integral. As mentioned before, this is in sharp contrast with the corresponding *single-source* versions of the MDP (polynomial-time solvable) and the UFP (*NP*-hard, but approximable to within $O(1)$ [31, 32, 15]). See Section 4 for further hardness results. Given this hardness, we formulate a new market-clearing mechanism; see Section 7. Informally, we consider Wt-(0/1-VERSION, MAX-#CONTRACTS) where, given a profit w_A for each $A \in R$, we want a $0/1$ solution that maximizes the profit of the fulfilled contracts. We assume by scaling that $w_A \in [0, 1]$ for all $A \in R$, and show a nearly best-possible bicriteria approximation. Let OPT be the optimum value of this problem. Given $\epsilon > 0$ and a flow f, let us say that the flow $(1 - \epsilon)$-fulfills contract $A \in R$ iff $f(A) \geq (1 - \epsilon) \cdot flow(A)$. Then, in polynomial time, we can find a flow in which the total profit of the $(1 - \epsilon)$-fulfilled contracts is at least: (i) $\Omega(\epsilon \cdot OPT^2/m)$ if $\epsilon \leq 1/2$, and (ii) $\Omega(OPT \cdot (OPT/m)^{(1-\epsilon)/\epsilon})$ if $\epsilon > 1/2$. (Note that if ϵ is "small", say 0.1, then we almost satisfy the demands of the $(1 - \epsilon)$-fulfilled contracts, while still remaining close to the $m^{1/2-\epsilon'}$-hardness-of-approximation result. If ϵ is larger, i.e., if we are willing to satisfy a smaller fraction of the demands, the objective function gets even better: in particular, if $\epsilon = 1 - \Theta(1/\log n)$, we get to within a constant factor of OPT. This suggests that when possible, we can choose such a relatively large ϵ and conduct the routing in *rounds*, where the routing is feasible in each round. Even if ϵ is $1 - \Theta(1/\log n)$, we require only $O(\log^2 n)$ rounds to fully satisfy the demands of the $(1 - \epsilon)$-fulfilled contracts.) The above assumes that $OPT < m$; we get even better results if $OPT \geq m$.

The above result follows from a more general multi-source, multi-sink result that we derive. As sketched in Section 3, multi-source, multi-sink problems are somewhat complicated by the fact that we cannot distin-

guish between the flows for different pairs (e.g., there may be no (s_i, t_i)-path in the flow graph); we show how this issue can be handled, and derive the single-source result as a corollary. For the multi-source, multi-sink *I-version* of the problem where we wish to maximize the total weighted flow, we build on the previous algorithm to deliver a solution of value $\Omega(OPT \cdot \min\{OPT, m\}/m)$. As mentioned in Section 8.2, a hardness result of [21] implies that for any fixed $\epsilon > 0$, approximating this problem to within $m^{1/2-\epsilon}$ is *NP*-hard; thus, our result for the I-version is essentially best-possible[1]. (The *I-version* is appropriate for Internet telephony/fax transmission where data is divided into atomic packets.) Work of [21] considers the relative of the *I-version* where we require that $f(A) \in \{0, flow(A)\}$ for all $A \in R$; it is also required that the flow of any commodity on any arc be an integer. Suppose all capacities are integers and that the maximum demand (i.e., maximum value of $flow(A)$) is d_{max}. An approximation guarantee of $O(\sqrt{md_{max}} \log^2 m)$ is presented for this problem, in [21].

Our approach is to conduct appropriate randomized rounding of multi-commodity flow relaxations as in [43], with two main new ideas. First, it is well-known that adding valid constraints to relaxations is often crucial for optimization/approximation. Our solutions depend on a key valid constraint (see (5)) added to the "natural" LP relaxation. Second, our bounds above for the case $\epsilon \leq 1/2$ depend on an existentially optimal tail probability bound that we derive–see Theorem 5.5; in its absence, our analysis would have only yielded the bound we get for $\epsilon > 1/2$ for the case $\epsilon \leq 1/2$. (Note that for any $OPT < m$, our first bound $\epsilon \cdot OPT^2/m$ is much higher than $OPT \cdot (OPT/m)^{(1-\epsilon)/\epsilon}$ as $\epsilon \to 0+$). This comes about by a careful analysis of the problem at hand, instead of straightforward application of a Chernoff-Hoeffding bound. This tail bound also improves the approximation for a class of packing integer programs due to [42]; see Theorem 8.1. Another aspect of our *I-version* algorithm is that it includes the design and analysis of a final "cleanup" phase to correct for some limits that may have been exceeded by the randomized rounding.

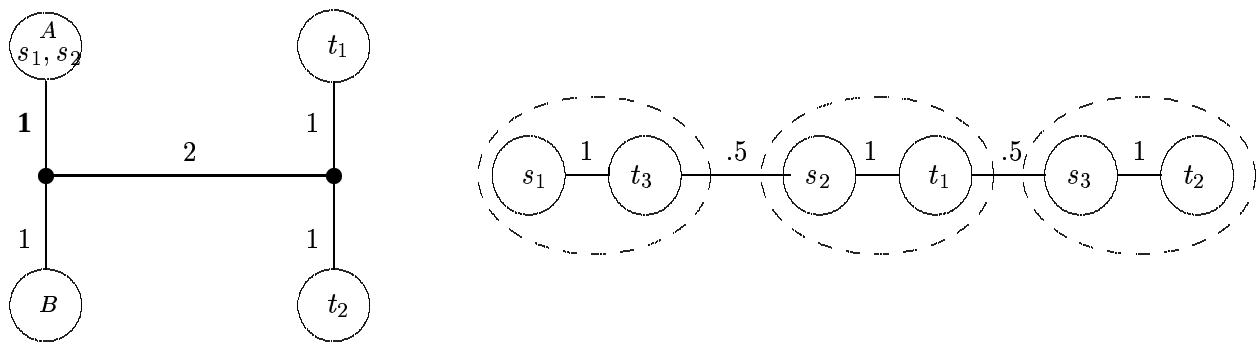
Three variants of our problems/methods have been studied by researchers: (i) Multi-commodity flow and related problems, (ii) unsplittable flow problems, and (iii) Lagrangean relaxation-type work in Operations Research. We briefly discuss two of these now. The basic Multi-commodity Flow problem consists of a network with capacities on the edges and a set of source-sink pairs $\{(s_1, t_1), \ldots, (s_k, t_k)\}$ with associated demands. We associate a commodity i with each source-sink pair (s_i, t_i). The problem is to find an integral flow that is a valid flow for each commodity such that the total flow on each edge does not exceed the capacity of that edge and all demands are satisfied. The problem is *NP*-hard and good approximations are known for certain optimization versions of the problem [38, 39]. Next, the unsplittable flow problem results from insisting that the flow for each (s_i, t_i), should be on a *single path* [30]; see, e.g., [30, 43, 33] for approximation algorithms for this problem. An important special case here is when all the source vertices s_i are the same, say s. Even here, it is *NP*-complete to decide if there is a *single* s–t_i path for each i such that the total demand using any edge does not exceed the edge's capacity. Constant-factor approximation algorithms are known for several optimization versions of this problem [30, 31, 32, 15], and near-optimal hardness-of-approximation results are known for its multi-source, multi-sink version [21].

3 Illustrative Examples: Structure of Solutions

We start by deriving some insights into the structure of solutions to the problems at hand. The examples will illustrate contrasts between this problem and related flow problems from the literature.

Example 1. This example illustrates the issues encountered as a result of deregulation. Figure 1(a) shows an example in which there are two power plants A and B. Let us say that each consumer has a demand of 1. Before deregulation, say both A and B are owned by the same company. If we assume that the plants

[1] The hardness result in [21] holds only for directed graphs.



1(a) **1(b)**

Figure 1: Figures for Examples 1 and 2.

have identical operating and production costs, then the demands can be satisfied by producing 1 unit of power at each plant. Now assume that due to deregulation, A and B are owned by separate companies. Further assume that A provides power at a much cheaper rate and thus both the consumers sign contracts with A. It is clear that both the consumers cannot now be provided power by A alone. Thus although the total production capacity available is more than total demand and it is possible to route that demand through the network under centralized control, it is not possible to route these demands in a deregulated scenario.

Example 2. Here, the graph consists of a simple line as shown in Figure 1(b). We have three contracts each with a demand of 1. The capacity of each edge is also 1. A feasible solution is $f(s_1, t_3) = f(s_2, t_1) = f(s_3, t_2) = 1$. The crucial point here is that *the flow originating at s_i may not go to t_i at all* — since power produced at the sources are indistinguishable, the flow from s_i joins a stream of other flows. If we look at the connected components induced by the edges with positive flow, we may have s_i and t_i in a different component. Thus we do not have a path or set of paths to round for the (s_i, t_i)-flow. This shows a basic difference between our problem and standard multi-commodity flow problems, and indicates that traditional rounding methods *may not be directly applicable.*

4 Hardness Results

4.1 Hardness of Simple Instances

We start by recalling two NP-hard problems. In the PARTITION problem, we are given a finite set $T = \{t_1, t_2, \ldots, \}$ of reals, and have to decide if the set can be partitioned into two subsets which sum up to the same value. In all our uses of this problem, we will let B denote $\sum_i t_i$. In the KNAPSACK problem, we are given a set S of items, each with a certain weight and a profit; given a budget W, we need to find a subset S' of S whose items add up to at most W in weight, such that the total profit of the elements in S' is maximized.

Theorem 4.1 *The* WT-(0/1-VERSION, MAX-FEASIBLE FLOW*) and* WT-(R-VERSION, MAX-FEASIBLE FLOW*) problems are NP-hard even for graphs with a single edge.*

Proof. Reduction from PARTITION. Given an instance $T = \{t_1, t_2, \cdots, t_n\}$ of PARTITION, we create a graph with two nodes u and v and one edge (u, v). The capacity of the edge (u, v) is $B/2$. There are n contracts; the ith contract has source node u and sink node v and the demand of the contract is t_i. It is now easy to see that there is a subset of contracts with flow value $B/2$ if and only if T has a solution. By a similar reduction, but

starting from KNAPSACK, we get the hardness of (0/1-VERSION, MAX-#CONTRACTS) with integral profit function. Note that (0/1-VERSION, MAX-#CONTRACTS) without profit function defined on a single edge is polynomial-time solvable by a simple greedy algorithm that chooses at each stage a contract with minimum demand. □

In our study, the problem of finding the best flows with fractional contracts but with certain costs on producing the commodity arose. If the cost function is linear, then the problem is easily seen to be in *P*, by using the standard flow formulation. But when we have non-linear costs for commodity production, then the problem turns out to be hard as shown below.

Theorem 4.2 *The* WT-*(*R-VERSION, MAX-FEASIBLE FLOW*)* , WT-*(*R-VERSION, MAX-#CONTRACTS*) problems subject to non-linear production costs and a budget on the total costs are NP-hard, even for graphs with a single edge.*

Proof. The reduction is almost identical to Theorem 4.1. We use non-linear costs to simulate 0/1-contracts; i.e., contracts are either chosen or not. The reduction is from the PARTITION problem. Given an instance $T = \{t_1, t_2, \ldots, t_n\}$ of the PARTITION problem, we create a graph with two nodes u and v and one edge (u, v). The capacity of the edge (u, v) is $B/2$. There are n contracts, one corresponding to each t_i. The ith contract has source node u and sink node v and the demand of the contract is t_i. The ith contract needs 0 dollars for producing 0 units at the source and requires t_i dollars for producing any amount above 0. It can produce no more than t_i units. It is easy to see here that once we choose a contract to produce power, we might as well ramp it up to its capacity; so, there is a subset of contracts with total production cost $B/2$ and flow value $B/2$ if and only if T has a solution. The reduction for (R-VERSION, MAX-#CONTRACTS) with profits is similar, except that we perform the reduction from KNAPSACK. □

We now consider the complexity of (0/1-VERSION, MAX-FEASIBLE FLOW) and (R-VERSION, MAX-FEASIBLE FLOW) with additional realistic constraints: *production constraints* on the amount of commodity that each producer can produce. We will employ the 3-PARTITION problem for this. An instance of 3-PARTITION consists of a set $T = \{t_1, \ldots t_{3m}\}$ such that each t_i has a weight $w_i \in [B/4, B/2]$, and such that $\sum_i w_i = mB$. The question is whether T can be partitioned into m subsets with 3 elements each, such the sum of the weights of the elements in each subset is exactly B.

Theorem 4.3 *(a) The (*0/1-VERSION, MAX-FEASIBLE FLOW*) problem is strongly NP-hard, when we place a budget on the amount of commodity that each producer can produce. (b) The (*0/1-VERSION, MAX-#CONTRACTS*) problem is NP-hard when we place a budget on the amount of commodity that each producer can produce.*

Proof. Starting from an instance I of 3-PARTITION, we create a star as follows. We have a center node u and spoke nodes s_1 to s_m each connected to the star node u. Each edge (u, s_i) has capacity B. The node u has $3m$ production stations $p_1, \ldots, p_{3m}$. Each spoke node s_i has $3m$ sinks named $t_1^i, \ldots, t_{3m}^i$. We have a contract (p_j, t_j^i) of value w_j for all (i, j), and each p_j has production capacity w_j. This implies that it can send power to only one of the requesting sinks at any given time, if we assume no fractional contracts. So we see that T has a solution if and only if there is a way to satisfy $3m$ contracts without violating the flow and capacity constraints. Thus the reduction works as an *NP*-hardness proof for the versions where (i) we want to satisfy the maximum number of contracts and (ii) we want to maximize the total flow in the network.

□

Note: All the above hardness results are for very simple graphs: each of these graphs is a tree and hence, simultaneously planar and of bounded treewidth.

4.2 Inapproximability Results

We strengthen the results in Section 4.1 and show that for arbitrary instances many of these problems are inapproximable: they do not even admit good approximation algorithms in general. Recall that in the maximum 3-DIMENSIONAL MATCHING (3DM) problem, we are given pairwise disjoint sets X, Y, Z, and a set $T \subseteq X \times Y \times Z$. The goal is to select a maximum-cardinality subset $T' \subseteq T$, such that the tuples in T' (viewed as 3-element sets) are pairwise disjoint. Also recall that in the INDEPENDENT SET problem on graphs, we are given an undirected graph, and wish to find an independent set (subset of the nodes in which no node is adjacent to any other node in the subset) of maximum cardinality. We start with a known result:

Theorem 4.4 ([1, 23, 40, 29]) *(i) Unless P=NP, $\exists \epsilon > 0$ such that no polynomial-time algorithm can guarantee an approximation factor of less than $(1+\epsilon)$, for* 3DM. *Moreover, there is a constant $B \geq 1$ such that this result holds even when each element in $X \cup Y \cup Z$ occurs in at most B tuples in T.*
(ii) Unless NP$\subseteq ZPP$, INDEPENDENT SET *does not have a polynomial-time approximation algorithm with performance guarantee $|V|^{1-\epsilon}$, where $|V|$ denotes the number of vertices in the input graph, and ϵ is an arbitrary positive constant.*

We begin with the case of bounded demands.

Theorem 4.5 *There is a constant B such that the following holds for (*0/1-VERSION, MAX-#CONTRACTS*), even when all edges have capacity* 1*, there is only one supplier, all vertices except the supplier have degree at most B, and all contracts have value* 3*. Unless P=NP, $\exists \epsilon > 0$ such that no polynomial-time algorithm can guarantee an approximation factor of less than $(1+\epsilon)$.*

Proof. We provide an approximation-preserving reduction from 3DM. Let the instance of 3DM consist of sets X, Y, Z and $T \subseteq X \times Y \times Z$. We construct an instance $G = (V \cup U, E)$ of the problem as follows. For each $t \in T$ we create a node $v_t \in V$ and for each $w \in X \cup Y \cup Z$ we create a node $u_w \in U$. We also have a supplier node $s \in U$. We add an edge with capacity 1 from s to each node in $U \setminus \{s\}$. For each $t = (x, y, z) \in T$, we add edges $(u_x, v_t), (u_y, v_t)$ and (u_z, v_t) with capacity 1. Each node in V has a contract with s with flow value 3. Thus if a contract v_t is completely satisfied, no other contract whose corresponding set in the 3DM instance has a non-empty intersection with t can also be completely satisfied. Therefore the optimal objective functions of the 3DM instance and the (0/1-VERSION, MAX-#CONTRACTS) instance are the same. We may now invoke Theorem 4.4(i) to complete the proof. ∎

Theorem 4.6 *The (*0/1-VERSION, MAX-#CONTRACTS*) problem is NP-hard even when restricted to instances G with the following constraints: (i) G is planar, (ii) each edge in G has capacity at most* 1*, (iii) each vertex has a bounded degree, and (iv) all contracts have value* 3*.*

Proof. We first modify the proof of Theorem 4.5 as follows: we have two identical copies of vertices V_1 and V_2 corresponding to T. The vertices $v_t^1 \in V_1$ and $v_t^2 \in V_2$ corresponding to a $t \in T$ will be called *twins*. We do not have the node s as in the proof of Theorem 4.5; instead, each node v_t^1 is now a supplier. For each $t = (x, y, z) \in T$, we add edges $(u_x, v_t^1), (u_y, v_t^1)$, (u_z, v_t^1), $(u_x, v_t^2), (u_y, v_t^2)$ and (u_z, v_t^2) with capacity 1. Each v_t^1 has a contract with its twin v_t^2 with flow value 3. Let $G = (V_1 \cup V_2 \cup U, E)$. It is once again easy to see that the optimal objective functions of the 3DM instance and this (0/1-VERSION, MAX-#CONTRACTS) instance are the same. Next, we modify the construction to obtain hardness for planar instances as follows.

Given an instance F of 3DM, its corresponding bipartite graph $BG(F)$ is defined naturally as follows: it is the same bipartite graph $G = (V \cup U, E)$ as in the proof of Theorem 4.5, with the vertex s removed. It

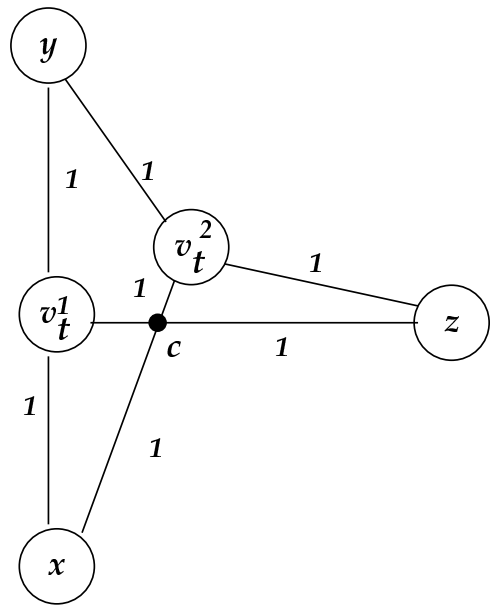


Figure 2: Schematic Diagram showing how to replace a crossover in the proof of Theorem 4.5 to obtain planar instances. The capacities of all edges are 1. The reduction locally replaces nodes in the bipartite graph representation by new nodes.

is known that 3DM is *NP*-hard even when restricted to instances F for which $BG(F)$ is planar [16]. Now, given such an F, we create an instance G of (0/1-VERSION, MAX-#CONTRACTS) as in the first part of the proof, and then modify it to make it planar as follows. Construct $BG(F)$, and lay each node in V_2 next to its twin in V_1. A typical such layout is depicted in Figure 2. This will introduce exactly one crossover per node in V_1 and as shown in Figure 2. We just replace the crossover by a new node $c = c_t$ and split the edges at that point. All the edges still have a capacity of 1. Let us see how to route 3 units of flow from v_t^1 to v_t^2. Since all edge capacities are 1, v_t^1 has to send exactly one unit on each of the edges (v_t^1, c), (v_t^1, x) and (v_t^1, y). The flow at x has to be routed through c and thus we have exactly 2 units of flow coming in at c and since there are 2 edges coming out, these can be sent out without violating any constraints.

The remaining details are similar to the proof of Theorem 4.5. □

Note that Theorem 4.5 provides inapproximability results restricted to non-planar instances with a single source. On the other hand, Theorem 4.6 only proves the *NP*-hardness of (0/1-VERSION, MAX-#CONTRACTS) for planar instances[2] but with multiple source-sink pairs. We now strengthen the results in Theorems 4.5 and 4.6 when demands grow polynomially with input size.

Theorem 4.7 *Unless NP $\subseteq ZPP$, no polynomial-time algorithm can guarantee a performance of $m^{\frac{1}{2}-\epsilon}$ for (*0/1-VERSION, MAX-#CONTRACTS*), for any fixed $\epsilon > 0$. This holds even when all edges have capacity* 1*, there is only one supplier node, and all contracts are integer-valued.*

Proof. We provide an approximation-preserving reduction from INDEPENDENT SET (IS) to the problem. If n and m are the number of nodes and edges in the IS instance, the number of arcs in the problem will be $O(n + m)$. Since Theorem 4.4(ii) shows that IS is also hard to approximate to within $m^{1/2-\epsilon}$, the theorem will follow. Let $H = (V, E)$ be the instance of IS. Create a graph $G' = (U \cup W, E')$ as follows. The set W

[2]The planar version of 3DM has a polynomial-time approximation scheme [2].

is a copy of V. We will abuse notation and refer to the copy of v in W as v. For every edge $e = \{u, v\} \in E$, create a node $x_e \in U$ and edges $\{x_e, u\}$ and $\{x_e, v\}$ in E'. Also create one "supply node" $s \in U$ and edges $\{\{s, x_e\} \mid e \in E\}$. All edges in G' have capacity 1. Each node in W has a contract with s with flow value equal to its degree in H (which is also its degree in G'). To satisfy the contract of any $w \in W$, there must be one unit of flow to w from each node in U adjacent to w. Since only one unit of flow can be sent to any node in $U \setminus \{s\}$, no other node in W adjacent to w in G can have its contract satisfied if w's contract is satisfied. So there is a bijection between feasible sets of contracts in G' and independent sets of the same size in H. □

Recall that, in contrast, such problems for single-source *unsplittable* flow have $O(1)$ [or $O(\log n)$ for the weighted case]-factor approximation algorithms. Our hardness result shows a significant difference between unsplittable flow and our splittable flow problems. One key cause of this difference is that (0/1-VERSION, MAX-#CONTRACTS) can specify in a very strict manner as to how the flow starting at a vertex can be split. For example, the (0/1-VERSION, MAX-#CONTRACTS) instance in the proof of Theorem 4.7 specifies that the flow to each node in W must be split along a number of paths equal to the degree of the node in H.

5 Useful probabilistic tools

Henceforth, an "efficient" algorithm will mean an algorithm running in polynomial time. We next present some (probabilistic) techniques relevant to our analysis. Let e denote the base of the natural logarithm. We abbreviate the phrase "random variable" by "r.v.".

Constructive versions of certain low-probability events. A significant result in derandomization techniques is that certain random structures that actually have very low probabilities of being generated by some underlying random process $\mathcal{R}$, can be constructed efficiently if $\mathcal{R}$ has some suitable structure. One such tool is Theorem 4.3 of [44], which appears as Theorem 5.2 here. To do so, we will need the following preliminaries from [44]. Given $\vec{a} = (a_1, a_2, \ldots, a_\ell) \in \{0,1\}^\ell$ and $\vec{b} = (b_1, b_2, \ldots, b_\ell) \in \{0,1\}^\ell$, let us say that $\vec{a} \preceq \vec{b}$ iff $a_i \leq b_i$ for all i. Suppose $X_1, X_2, \ldots, X_\ell$ are *independent* r.v.s, each taking values in $\{0,1\}$. Let $\vec{X} \doteq (X_1, X_2, \ldots, X_\ell)$. Define an event $\mathcal{A}$ to be *increasing* iff: for all $\vec{a} \in \{0,1\}^\ell$ such that $\mathcal{A}$ holds when $\vec{X} = \vec{a}$, $\mathcal{A}$ also holds when $\vec{X} = \vec{b}$, for any $\vec{b}$ such that $\vec{a} \preceq \vec{b}$. Analogously, event $\mathcal{A}$ is said to be *decreasing* iff: for all $\vec{a} \in \{0,1\}^\ell$ such that $\mathcal{A}$ holds when $\vec{X} = \vec{a}$, $\mathcal{A}$ also holds when $\vec{X} = \vec{b}$, for any $\vec{b} \preceq \vec{a}$.

The basic way in which Theorem 5.2 will be useful for us is as follows. Suppose $E_1, E_2, \ldots, E_t$ are some events that are determined completely by $\vec{X}$. Each E_i is a "bad" event from our perspective; we want to efficiently find a value for the vector $\vec{X}$ under which *none* of the events E_i hold. Theorem 5.2 presents a useful sufficient condition to this end; we next present some further notation in order to understand Theorem 5.2. An r.v. g is said to be a *well-behaved* estimator for an event $\mathcal{E}$ (w.r.t. $\vec{X}$) iff it satisfies the following properties (P1), (P2), (P3) and (P4), $\forall u \leq \ell$, $\forall T = \{i_1, i_2, \ldots, i_u\} \subseteq [\ell]$, and for all $b_1, b_2, \ldots, b_u \in \{0,1\}$ for which $\Pr(\bigwedge_{s=1}^{u}(X_{i_s} = b_s)) = \prod_{s=1}^{u} \Pr(X_{i_s} = b_s) > 0$. For notational convenience, we let $\mathcal{B}$ denote "$\bigwedge_{s=1}^{u}(X_{i_s} = b_s)$".

(P1) $\mathbf{E}[g|\mathcal{B}]$ is efficiently computable: i.e., computable in $poly(\ell)$ time;

(P2) $\Pr(\mathcal{E}|\mathcal{B}) \leq \mathbf{E}[g|\mathcal{B}]$;

(P3) *if $\mathcal{E}$ is increasing*, then for all $i_{u+1} \in ([\ell] - T)$ with $\Pr(X_{i_{u+1}} = 1) \in (0,1)$, $\mathbf{E}[g|(X_{i_{u+1}} = 0) \wedge \mathcal{B}] \leq \mathbf{E}[g|(X_{i_{u+1}} = 1) \wedge \mathcal{B}]$; and

(P4) *if $\mathcal{E}$ is decreasing*, then for all $i_{u+1} \in ([\ell] - T)$ with $\Pr(X_{i_{u+1}} = 1) \in (0,1)$, $\mathbf{E}[g|(X_{i_{u+1}} = 1) \wedge \mathcal{B}] \leq \mathbf{E}[g|(X_{i_{u+1}} = 0) \wedge \mathcal{B}]$.

Remark 5.1 The condition "$\Pr(X_{i_{u+1}} = 1) \in (0,1)$" is specified in (P3) and (P4) for the following reason: if $\Pr(X_{i_{u+1}} = 1) \in \{0,1\}$, then one of the two terms "$\mathbf{E}[g|(X_{i_{u+1}} = 0) \wedge \mathcal{B}]$" and "$\mathbf{E}[g|(X_{i_{u+1}} = 1) \wedge \mathcal{B}]$" will be undefined. To gain some intuition about (P1)–(P4), note that if g is $\chi(\mathcal{E})$, the indicator r.v. for $\mathcal{E}$, it satisfies (P2), (P3) and (P4). However, in our applications, it will be unclear if (P1) is true if $g = \chi(\mathcal{E})$. Thus, we seek a r.v. g which "mimics" $\chi(\mathcal{E})$ in that (P2), (P3) and (P4) hold; we also want the "efficient computability" property (P1) to hold for g.

If g satisfies (P1) and (P2) (but not necessarily (P3) and (P4)), we call it a *proper* estimator for $\mathcal{E}$ w.r.t. $\vec{X}$. For any r.v. X and event $\mathcal{A}$, let $\mathbf{E'}[X]$ denote $\min\{\mathbf{E}[X], 1\}$.

Theorem 5.2 ([44]) *Suppose $\vec{X} = (X_1, X_2, \ldots, X_\ell)$ is a sequence of* independent *r.v.s X_i, with $X_i \in \{0,1\}$ for each i. Let $E_1, E_2, \ldots, E_t$ be events and r, s be non-negative integers with $r + s \leq t$ such that:*
(i) $E_1, E_2, \ldots, E_r$ are all increasing*, with respective well-behaved estimators $g_1, g_2, \ldots, g_r$ w.r.t. $\vec{X}$;*
(ii) $E_{r+1}, \ldots, E_{r+s}$ are all decreasing*, with respective well-behaved estimators $g_{r+1}, \ldots, g_{r+s}$ w.r.t. $\vec{X}$;*
(iii) $E_{r+s+1}, \ldots, E_t$ are arbitrary events, with respective proper estimators $g_{r+s+1}, \ldots, g_t$, and
(iv) all the E_i and g_i are completely determined by $\vec{X}$.

Then, if

$$1 - (\prod_{i=1}^{r}(1 - \mathbf{E'}[g_i])) + 1 - (\prod_{i=r+1}^{r+s}(1 - \mathbf{E'}[g_i])) + \sum_{i=r+s+1}^{t} \mathbf{E}[g_i] < 1 \tag{1}$$

holds, we can construct a value for $\vec{X}$ in poly$(\ell + t)$ time deterministically, under which none of the E_i hold. (As usual, empty products are taken to be 1; e.g., if $s = 0$, then $\prod_{i=r+1}^{r+s}(1 - \mathbf{E'}[g_i]) \equiv 1$.)

The main point is that $\Pr(\bigwedge_i \overline{\mathcal{E}_i})$ can be tiny: say, exponentially small in $(\ell + t)$. However, as long as the requirements of Theorem 5.2 are met, we can efficiently set values for the X_i so that all the $\mathcal{E}_j$ are avoided.

Remark 5.3 As pointed out in [44], there is a simple way of checking properties (P1)–(P4) above, for a given r.v. g. Recall that in our definition of well-behavedness, $\mathcal{B}$ is short-hand for "$\bigwedge_{s=1}^{u}(X_{i_s} = b_s)$", where the indices i_s as well as the bits b_s are arbitrary. Now, even conditional on $\mathcal{B}$, the X_i can be viewed as *independent* r.v.s: it is only that for each i_s, $X_{i_s} = b_s$ with probability 1. Thus, in order to check any of the properties (P1)–(P4), it suffices to check that for *any choice of* the probabilities $\Pr(X_i = 1),\ i = 1, 2, \ldots, \ell$, the properties hold when $u = 0$ (i.e., when $\mathcal{B}$ is the tautology).

A new tail inequality. We now present an improved tail bound for a certain problem, in Theorem 5.5. This is a result where more work is required than a straightforward application of Theorem 5.4. We start by recalling the standard Chernoff-Hoeffding bounds.

Theorem 5.4 (Chernoff-Hoeffding bounds [8, 27]) *Let $X_1, X_2, \ldots, X_\ell$ be* independent *r.v.s, each taking values in $[0,1]$, with $R = \sum_{i=1}^{\ell} X_i$ and $\mathbf{E}[R] \leq \mu$. For any $\delta \geq 0$,*

$$\Pr(R \geq \mu(1+\delta)) \leq \mathbf{E}[(1+\delta)^{R-\mu(1+\delta)}] \leq G(\mu, \delta) \doteq (e^{\delta}/(1+\delta)^{(1+\delta)})^{\mu}; \tag{2}$$

if $\mathbf{E}[R] \geq \mu$ and $0 \leq \delta < 1$,

$$\Pr(R \leq \mu(1-\delta)) \leq \mathbf{E}[(1-\delta)^{R-\mu(1-\delta)}] \leq H(\mu, \delta) \doteq e^{-\mu\delta^2/2}. \quad \blacksquare \tag{3}$$

Theorem 5.5 *Let $\vec{X} = (X_1, X_2, \ldots, X_\ell)$ be a sequence of ℓ* independent *random variables each taking values in $\{0,1\}$, and let $Z = \sum_i r_i X_i$, where each r_i lies in $[0,1]$. Suppose $\mathbf{E}[Z] \leq 1/\lambda$ for some $\lambda > 1$. Then, for any $\epsilon > 0$, there is an explicitly presented well-behaved estimator g for the (increasing) event "$\Pr(Z \geq (1+\epsilon))$" w.r.t. $\vec{X}$; also, $\mathbf{E}[g] \leq (e^2 + \max\{2, \epsilon^{-1}\})/\lambda^2$.*

The proof of Theorem 5.5 is shown in the appendix. To appreciate the theorem, think of $\lambda > 1$ as "large" and ϵ as a "small" positive constant such as 0.2; we want a tail bound, i.e., a bound on $\Pr(Z \geq (1+\epsilon))$, that is $O(\epsilon^{-1}\lambda^{-2})$. Direct use of Markov's inequality and of (2) only yields the bounds $O(1/\lambda)$ and $O(\lambda^{-(1+\epsilon)})$ respectively. Also, this $O(\epsilon^{-1}\lambda^{-2})$ bound is existentially optimal to within a constant factor, for all $\lambda \geq (2\epsilon)^{-1}$. To see this, suppose $\ell = 2$, $r_1 = \epsilon$, $r_2 = 1$, $\mathbf{E}[X_1] = (2\epsilon\lambda)^{-1}$, and $\mathbf{E}[X_2] = (2\lambda)^{-1}$. Then, $\Pr(Z \geq (1+\epsilon)) = \Pr(X_1 = X_2 = 1) = (4\epsilon\lambda^2)^{-1}$.

6 New market-clearing mechanism

We start by describing a multi-source, multi-sink problem that can be viewed as a new market-clearing mechanism. As mentioned earlier, one of the main corollaries of our approximation algorithm for this problem, is an approximation algorithm for SINGLE SOURCE-(0/1-VERSION, MAX-#CONTRACTS). We start with an informal description, and then formally present the multi-source, multi-sink flow problem through constraints (A1)–(A4) below. By a *non-negative flow* f in a digraph $G = (V, E)$, our present context will simply mean a set of flow values $\{f_{u,v} \geq 0 : (u,v) \in E\}$; these values need not satisfy any conservation constraints etc., unless otherwise specified. Given f, the *net outflow* from $v \in V$, $f_{out}(v)$, is defined naturally to be $(\sum_{u:\ (v,u)\in E} f_{v,u}) - (\sum_{u:\ (u,v)\in E} f_{u,v})$. Similarly, the net inflow into v is $f_{in}(v) = -f_{out}(v)$. For any non-negative integer k, let $[k] \doteq \{1, 2, \ldots, k\}$.

We are given a digraph $G = (V, E)$ and *disjoint* subsets S (for "suppliers") and C (for "customers") of V. (If G is undirected, replace each edge $\{u, v\}$ by the arcs (u, v) and (v, u).) Each edge $(u, v) \in E$ has a *capacity* $c_{u,v} > 0$; each $s \in S$ has a *cost* $p(s) > 0$, and each $t \in C$ has a *demand* $d_t > 0$ and a *weight* $w_t > 0$. (By scaling, we will assume that the costs and weights lie in $[0, 1]$.) We are also given a *budget* B. We will denote $|V| = n$ and $|E| = m$ throughout. Each $s \in S$ is a supplier who charges $p(s)$ per unit outflow from s; the weight w_t for $t \in C$ reflects the "importance" of customer t. We wish to construct a non-negative flow f in G and to choose an $A \subseteq C$, such that: (i) if $v \notin S$, then $f_{out}(v) \leq 0$ (i.e., no vertex outside of S can send out net positive flow) and if $v \notin C$, then $f_{in}(v) \leq 0$; (ii) for all $t \in A$, there is a net inflow of at least d_t (or, in a slightly relaxed setting, at least $d_t(1-\epsilon)$ for some given $\epsilon > 0$), and (iii) no edge carries a flow more than its capacity.

We next discuss payment policy. Given f, note that $f_{out}(s)$ for any $s \in S$ can be viewed as the amount of flow contributed by s. Suppose s charges $p(s)$ per unit net outflow. Then, we can see that $\sum_{s\in S} f_{out}(s) = \sum_{t\in C} f_{in}(t)$, which equals some F, say. Thus, supplier s provides a fraction $f_{out}(s)/F$ of the total flow. Hence, each $t \in C$ can pay $f_{in}(t) \cdot p(s) \cdot f_{out}(s)/F$ to each supplier s. This way, each customer t pays an appropriate fraction to each supplier (i.e., divides up his payment in an appropriate way among the suppliers), and each supplier gets her rightful total amount of $(\sum_{t\in C} f_{in}(t) \cdot p(s) \cdot f_{out}(s)/F) = p(s) f_{out}(s)$. Thus, the total amount paid is $\sum_{s\in S} p(s) f_{out}(s)$, and our next constraint is that this should not exceed the budget B: (iv) $\sum_{s\in S} p(s) f_{out}(s) \leq B$. Finally, the objective is to maximize $\sum_{t\in A} w_t$ (the weighted number of satisfied customers).

Formally, the problem $\mathcal{P}$ is to construct a non-negative flow f in G and to choose an $A \subseteq C$, such that:

(A1) $\forall v \notin S$, $f_{out}(v) \leq 0$ and $\forall v \notin C$, $f_{in}(v) \leq 0$;

(A2) $\forall t \in A$, $f_{in}(t) \geq d_t$;

(A3) $\forall (u,v) \in E$, $f_{u,v} \leq c_{u,v}$, and

(A4) $\sum_{s \in S} p(s) f_{out}(s) \leq B$.

Subject to these constraints, we wish to maximize $\sum_{t \in A} w_t$. We shall mainly focus on the "ϵ-relaxed" variant of $\mathcal{P}$ wherein condition (A2) is weakened to "(A2'): $\forall t \in A,\ f_{in}(t) \geq d_t(1-\epsilon)$". Setting $B = \infty$ in the ϵ-relaxed version of $\mathcal{P}$ yields the ϵ-relaxed version of a common generalization of (0/1-VERSION, MAX-#CONTRACTS) and 0/1-MAX-FLOW where, given a profit w_A for each $A \in R$, we want a $0/1$ solution that maximizes the profit of fulfilled contracts. Theorem 7.2 will present an approximation algorithm for the ϵ-relaxed version of $\mathcal{P}$.

6.1 Rationale for the Model

Although the model proposed above relaxes the notion of contracts as defined earlier, the relaxed version is directly motivated by the *Poolco model* currently employed in California [7]. In this model, each producer bids for various amounts of power that it is willing to sell at particular prices. The customers do not bid for power; the ISO uses simple mechanisms to "heuristically" select the cheapest set of producers who can fulfill the requirement. The main difference though is that the strike price of power chosen is the same across all producers and in general this is equal to the highest bid price at which the market clears (this implies balancing the supply and demand while obeying the network capacity constraints). The proposed model tries to overcome the obvious drawback in the current model used in California (referred as the Cal-ISO model); namely by trying to minimize the total cost incurred by the cunsomers (specified by a budget B). We believe that the proposed model is much better in terms of avoiding escalating power prices that are currently witnessed in the California power market (see August 04 and August 14 2000 articles in the *Wall Street Journal* [48] that elaborate on this subject).

7 Approximation algorithms for the new market-clearing mechanism

In this section, we present approximation algorithms for a certain multi-source, multi-sink flow problem that is presented in Section 6, and derive an algorithm for SINGLE SOURCE-(0/1-VERSION, MAX-#CONTRACTS) as a corollary. The algorithms start by modeling the problem as an Integer Linear Program (ILP). The integrality constraints of the ILP are relaxed, leading to a linear program (LP); the LP can be solved efficiently using, e.g., well-known polynomial time algorithms for linear programming. One of the noteworthy points here is the addition of certain *valid constraints* to the LP. We then present a randomized rounding approach to round the fractional solution obtained, and show how to extract an efficient deterministic rounding algorithm from it with provable worst-case guarantees. This overall scheme for devising approximation algorithms above has been studied previously by many researchers. The novel aspects of our work are the introduction of certain valid constraints, as well the existentially optimal tail probability bound of Theorem 5.5. These lead to near-optimal performance bounds.

7.1 An LP relaxation and its rounding

A good (linear programming) relaxation for $\mathcal{P}$ or for $\mathcal{P}$'s ϵ-relaxation, may not be immediate because of flow-indistinguishability; a little care helps. Suppose we have a non-negative flow f in a digraph $G = (V, E)$, such that for some given $h(\cdot)$,

$$\forall (u,v) \in E,\ f_{u,v} \leq c_{u,v},\ \text{and}\ \forall v \in V,\ f_{out}(v) = h(v). \tag{4}$$

Via standard "flow decomposition" ideas (see [3]), f can be efficiently transformed into a set of "path flows" that still essentially satisfy (4) in the following sense. We can efficiently construct a set of flow-paths $\{P_i : 1 \leq i \leq \ell\}$, where P_i is from some u_i to some v_i and has some flow value $a_i > 0$. We can ensure the following properties.

(B0) $\ell \leq m$;

(B1) $\forall v \in V$:

- (i) If $h(v) > 0$, then $v \neq v_i$ for any $i \in [\ell]$; also, $\sum_{i:v=u_i} a_i = h(v)$;
- (ii) if $h(v) < 0$, then $v \neq u_i$ for any $i \in [\ell]$; also, $\sum_{i:v=v_i} a_i = -h(v)$;
- (iii) if $h(v) = 0$, then $v \neq u_i$ and $v \neq v_i$ for any $i \in [\ell]$.

(B2) $\forall (u,v) \in E$, $\sum_{i:(u,v)\in P_i} a_i \leq c_{u,v}$.

Consider the following LP, (LP1), related to our main problem. Let $C = \{t_1, t_2, \ldots, t_k\}$. The LP is to define k real variables $x_1, \ldots, x_k$ each lying in $[0,1]$, and to construct k non-negative flows $f^{(1)}, f^{(2)}, \ldots f^{(k)}$ on G such that:

(C1) in each flow $f^{(i)}$, flow-conservation is satisfied at all nodes in $V - (S \cup \{t_i\})$, t_i is the only node allowed to have a positive in-flow, and $f^{(i)}_{in}(t_i) \geq d_i x_i$;

(C2) the total flow on any edge (u,v) is at most $c_{u,v}$, i.e., $\sum_{i\in[k]} f^{(i)}_{u,v} \leq c_{u,v}$;

(C3) $(\sum_{s\in S} p(s)(\sum_{i\in[k]} f^{(i)}_{out}(s))) \leq B$, and, crucially,

(C4) $$\forall (u,v) \in E \; \forall i \in [k], \; f^{(i)}_{u,v} \leq c_{u,v} x_i. \tag{5}$$

Subject to these constraints, the objective is to maximize $\sum_{i\in[k]} w_i x_i$.

It is easy to check that the above can be written as an LP, and we now show that any optimal integral solution to our problem leads to a feasible solution to this LP. Given any optimal integral solution, define $x_i = 1$ if $i \in A$, and 0 otherwise. Do a flow-decomposition, and note from (B1) that only vertices in S can be the sources in the resulting flow paths, and that only vertices in A will be sinks. Now interpret the set of all the flow-paths that end in any given $t_i \in A$ as the flow $f^{(i)}$ for our LP. From (B2), we also see that the total flow on any edge (u,v) is at most $c_{u,v}$; the budget constraint of our LP (constraint (C3)) is also satisfied. Constraint (C4) will be crucial; we now check that (C4) is also satisfied (i.e., that it is a valid constraint). If $i \notin A$ in the given optimal integral solution, then note that no net positive in-flow will be sent to t_i in this integral solution; thus, if $x_i = 0$, then (C4) holds. On the other hand, if the given optimal integral solution sets $i \in A$, then we have $x_i = 1$ and (C4) holds since (C2) does. Thus, any optimal integral solution leads to a feasible solution to (LP1) with the same objective function value; hence we indeed have a relaxation. Let OPT and y^* respectively denote the optimal values for our problem and for (LP1); we have as usual that $OPT \leq y^*$.

We now show how to round an optimal solution to (LP1), to solve the "ϵ-relaxed" variant of our problem. Start with an optimal solution to (LP1), and conduct a flow decomposition. For each $t_i \in C$, we get a set of flow-paths $P_{i,1}, P_{i,2}, \ldots$, each $P_{i,j}$ originating from some $s_{i,j} \in S$ and carrying a flow of value $z_{i,j} \geq 0$; (C1) shows that $\sum_j z_{i,j} = d_i x_i$. Let $\gamma > 1$ be a parameter that will be chosen below. Independently for each i, set a random variable Y_i to 1 with probability x_i/γ, and $Y_i := 0$ with probability $1 - x_i/\gamma$. If $Y_i = 1$, we will choose to satisfy an $(1-\epsilon)$-fraction of t_i's demand; i.e., for all j, we will multiply the flow values $z_{i,j}$ by

$(1-\epsilon)/x_i$. If $Y_i = 0$, we will choose to have zero flow sent to t_i: i.e., we will reset all the $z_{i,j}$ to 0. This flow yields our final flow, and we set $A = \{t_i \in C : Y_i = 1\}$. We now analyze this rounding process and also select a suitable $\gamma > 1$.

7.2 Analysis of the Rounding

Recall that m denotes the number of edges in G. We start by listing the $m+2$ events that we wish to avoid simultaneously:
(a) For any edge $(u,v) \in E$, let $E_{u,v}$ be the "bad" event that the final total flow $f_{u,v}$ on it is more than $c_{u,v}$. We have m such bad events.
(b) Let TP be the r.v. denoting the final payment; let E' be the bad event that $TP > B$.
(c) Recall that the objective function OBJ equals $\sum_{i\in[k]} w_i Y_i$. Our final bad event E'' is that $OBJ \leq y^*/(2\gamma)$.

We now show how to choose a suitable $\gamma > 1$ and then apply Theorem 5.2 in order to avoid all of the above $m+2$ events. As a result, we will obtain a solution to the ϵ-relaxed variant of our problem, with objective function value $\Omega(y^*/\gamma)$. The underlying independent binary random variables in the sense of Theorem 5.2, are now the r.v.s $Y_1, Y_2, \ldots, Y_k$. Further note that the first $m+1$ events above (i.e., those of (a) and (b)) are increasing, while the last event E'' is decreasing. In order to employ Theorem 5.2, we need well-behaved estimators for our $m+2$ bad events; this is simple for events E' and E''. First, it is easy to verify that $g' = TP/B$ is a well-behaved estimator for E'. Second, since $OBJ = \sum_{i\in[k]} w_i Y_i$ has mean y^*/γ, we can check using (3) that

$$g'' = 2^{y^*/(2\gamma) - OBJ} = 2^{y^*/(2\gamma)} \cdot \prod_{i\in[k]} 2^{-w_i Y_i}$$

is a well-behaved estimator for E''. Also, due to the scaling down by γ, we have $\mathbf{E}[TP] \leq B/\gamma$; this and (3) show respectively that

$$\mathbf{E}[g'] \leq 1/\gamma; \;\; \mathbf{E}[g''] \leq e^{-y^*/(8\gamma)}. \tag{6}$$

We finally consider the m events in (a). Consider the bad event $E_{u,v}$. We can see that $f_{u,v} = \sum_{(i,j):(u,v)\in P_{i,j}} (z_{i,j}(1-\epsilon)/x_i) Y_i$; hence,

$$\mathbf{E}[f_{u,v}] = \sum_{(i,j):(u,v)\in P_{i,j}} (z_{i,j}(1-\epsilon)/x_i) \cdot (x_i/\gamma) \leq c_{u,v}(1-\epsilon)/\gamma, \tag{7}$$

since $\sum_{(i,j):(u,v)\in P_{i,j}} z_{i,j} \leq c_{u,v}$ by (C2) and (B2). We have

$$E_{u,v} \equiv (\sum_{(i,j):(u,v)\in P_{i,j}} \frac{z_{i,j}}{x_i c_{u,v}} Y_i > 1 + \frac{\epsilon}{1-\epsilon}). \tag{8}$$

Crucially, we can deduce from (C4) that for all (i,j) and all edges (u,v) in $P_{i,j}$, $z_{i,j} \leq x_i c_{u,v}$. Also, (7) shows that

$$\mathbf{E}[\sum_{(i,j):(u,v)\in P_{i,j}} \frac{z_{i,j}}{x_i c_{u,v}} Y_i] \leq 1/\gamma.$$

Thus, by Theorem 5.5, there is a well-behaved estimator $g_{u,v}$ for $E_{u,v}$, with

$$\mathbf{E}[g_{u,v}] \leq (e^2 + \max\{2, (1-\epsilon)\epsilon^{-1}\})/\gamma^2. \tag{9}$$

There is a better alternative to this choice of $g_{u,v}$, if $\epsilon > 1/2$. The bound (2) shows that

$$g_{u,v} \doteq (1-\epsilon)^{(1-\epsilon)^{-1} - \sum_{(i,j):(u,v)\in P_{i,j}} z_{i,j}/(x_i c_{u,v}) Y_i}$$

is also a suitable well-behaved estimator for $E_{u,v}$, with

$$\mathbf{E}[g_{u,v}] \leq (\frac{e(1-\epsilon)}{\gamma})^{(1-\epsilon)^{-1}}. \tag{10}$$

Let p be the bound on $\mathbf{E}[g_{u,v}]$ that we choose, either from (9) or from (10). In order to simultaneously avoid all the m events $E_{u,v}$ and the events E' and E'', Theorem 5.2 and (6) show that the following condition is sufficient:

$$(1-1/\gamma)\cdot(1-\min\{p,1\})^m > e^{-y^*/(8\gamma)}. \tag{11}$$

Remark 7.1 Here and a few times more in the paper, we will need to choose the scale factor $\gamma > 1$ to satisfy bounds such as (11). We will always choose $\gamma \geq 2$; also, we will ensure that γ is large enough so that the term "p" is at most $1/2$. Recall that for $0 \leq x \leq 1/2$, $1-x \geq e^{-2x}$. Thus, e.g., in (11), it will suffice to show that $p \leq 1/2$, and that $2/\gamma + 2mp < y^*/(8\gamma)$. This recipe of ensuring that $p \leq 1/2$ and lower-bounding a term such as $(1-p)^m$ by e^{-2mp}, will be useful.

Following Remark 7.1, a simple calculation shows that if constants K_0, K_1 and K_2 are chosen large enough, (11) will hold in each of the following three cases. (In (i) and (ii), we use the bound (9) on $\mathbf{E}[g_{u,v}]$; in (iii), we use the bound (10) on $\mathbf{E}[g_{u,v}]$.) (i) $\gamma = K_0/\sqrt{\epsilon}$, if $y^* \geq m$; (ii) $\gamma = K_1 m/(\epsilon y^*)$, if $y^* < m$ and $0 < \epsilon \leq 1/2$; (iii) $\gamma = K_2(m/y^*)^{(1-\epsilon)/\epsilon}$, if $y^* < m$ and $1/2 < \epsilon \leq 1$. We thus get

Theorem 7.2 *Consider the multi-source, multi-sink flow problem $\mathcal{P}$. There are positive constants C_0, C_1 and C_2 such that for the ϵ-relaxed version of problem $\mathcal{P}$, we can output a feasible solution of the following value in deterministic polynomial time. (i) $C_0\sqrt{\epsilon}y^*$, if $y^* \geq m$; (ii) $C_1\epsilon(y^*)^2/m$, if $y^* < m$ and $0 < \epsilon \leq 1/2$; (iii) $C_2 y^* \cdot (y^*/m)^{(1-\epsilon)/\epsilon}\}$, if $y^* < m$ and $1/2 < \epsilon \leq 1$.*

In particular, part (ii) provides a bicriteria result for the general problem $\mathcal{P}$, that is not far from the hardness threshold for the single-source case.

8 Extensions and Applications

We now present some extensions and applications of the problems and ideas of Section 7.

8.1 Approximating packing integer programs

An application of Theorem 5.5 is to improving certain approximation bounds shown in [42] for a family of packing integer programs (PIPs). Given $A \in [0,1]^{m\times n}$, $b \in [0,\infty)^m$ and $c \in [0,1]^n$ with $\max_j c_j = 1$, a PIP seeks to maximize $c^T \cdot x$ subject to $x \in Z_+^n$ and $Ax \leq b$. We also define $B = \min_i b_i$; we can take $B \geq 1$ without loss of generality [42]. PIPs model various *NP*-hard problems such as knapsack, independent sets in graphs, and matchings in hypergraphs. A natural LP relaxation for a PIP is to relax "$x \in Z_+^n$" to "$x \in \Re_+^n$". (We could also have constraints such as "$x_j \in \{0,1,\ldots,d_j\}$", for some given integers d_j. In this case, the LP relaxation lets x_j be a real lying in $[0,d_j]$.) The current-best approximation bound for general PIPs, due to [42], is as follows: in deterministic polynomial-time, we can output a solution of value $\Omega(\min\{y^*, (K_3 y^*/m^{1/B})^{B/(B-1)}\})$, where $K_3 \in (0,1)$ is an absolute constant. (If $B = 1$, then a solution of

value $\Omega(y^*/m)$ can be output.) One problem with this result is that it becomes rather weak as B approaches 1 from above. We improve this result of [42] for the range $B \in (1,2)$, in Theorem 8.1. (Think of B as one plus a "small" constant such as 0.1.)

Theorem 8.1 *Given a general packing integer program with $B \in (1,2)$, we can in polynomial time output a solution of value $\Omega((B-1)y^* \cdot \min\{y^*, m\}/m)$.*

Proof. We multiply each row i of the linear system "$Ax \leq b$" by B/b_i, so that all entries in the vector b become B; the entries of A still lie in $[0,1]$. We follow the approach of [42]; the only difference is that we now employ Theorem 5.5 instead of directly plugging in the Chernoff-Hoeffding bounds. Solve the LP relaxation; let x_j^* be the value for x_j in the LP's computed optimal solution. Let $\gamma = (K'B/(B-1)) \cdot \max\{m/y^*, 1\}$, where $K' \geq 2$ is a sufficiently large constant. Independently for each j, we set $X_j = \lfloor x_j^*/\gamma \rfloor$ with probability $\lceil x_j^*/\gamma \rceil - x_j^*/\gamma$; we set $X_j = \lceil x_j^*/\gamma \rceil$ with probability $1 - (\lceil x_j^*/\gamma \rceil - x_j^*/\gamma)$. This is our rounding; we wish to show that if K' is large enough, then we can satisfy all the constraints and have the objective function value being at least $y^*/(2\gamma)$.

Define $z_j = X_j - \lfloor x_j^*/\gamma \rfloor$; note that $z_j \in \{0,1\}$. For each $i \in [m]$, let $s_i = \sum_j A_{i,j} \lfloor x_j^*/\gamma \rfloor$, and $b'_i = B - s_i$. Finally, let $y_0^* = \sum_j c_j \lfloor x_j^*/\gamma \rfloor$. If $y_0^* \geq y^*/(2\gamma)$, rounding down each x_j^*/γ will give us a feasible integral solution of value at least $y^*/(2\gamma)$. So we assume that $y_0^* < y^*/(2\gamma)$. It suffices to avoid the following $m+1$ events: (a) events $E_1, E_2, \ldots, E_m$, where $E_i \equiv ((Az)_i > b'_i)$, and (b) event $E_{m+1} \equiv (c^T \cdot z < (y^*/(2\gamma) - y_0^*))$. Consider any $i \in [m]$; note that $\mu_i \doteq \mathbf{E}[(Az)_i] \leq B/\gamma - s_i$. If $s_i = B/\gamma$, event E_i cannot happen; so assume $s_i < B/\gamma$. Now, $E_i \equiv ((Az)_i > B - s_i)$. Recall that $\gamma \geq 2B/(B-1)$, and that $s_i < B/\gamma$. Thus, $B - s_i > B - (B-1)/2 = (B+1)/2 > 1$. We now apply Theorem 5.5 to the event $((Az)_i > B - s_i)$; in the notation of Theorem 5.5, we may take $1/\lambda = B/\gamma - s_i$, and $\epsilon = B - s_i - 1$. By Theorem 5.5, we have a well-behaved estimator g_i for the event E_i, with $\mathbf{E}[g_i] \leq (e^2 + \max\{2, \epsilon^{-1}\})/\lambda^2$. Given our values of ϵ and λ and the fact that $B > B/\gamma + 1$, it is easy to check that in the range $s_i \in [0, B/\gamma)$, this bound on $\mathbf{E}[g_i]$ is maximized when $s_i = 0$. Thus,

$$\forall i \in [m],\ \mathbf{E}[g_i] \leq (B/\gamma)^2 \cdot (e^2 + \max\{2, (B-1)^{-1}\}). \tag{12}$$

We next consider the objective function. Note that $\mu_{m+1} \doteq \mathbf{E}[c^T \cdot z] = y^*/\gamma - y_0^*$, and define $\delta_{m+1} \in (0,1)$ by $\mu_{m+1}(1 - \delta_{m+1}) = y^*/(2\gamma) - y_0^*$; we have $E_{m+1} \equiv (c^T \cdot z < \mu_{m+1}(1 - \delta_{m+1}))$. By (3), there is a well-behaved estimator

$$g_{m+1} \doteq (1 - \delta_{m+1})^{y_0^* - y^*/(2\gamma)} \cdot \prod_j (1 - \delta_{m+1})^{c_j z_j}$$

for E_{m+1}, with $\mathbf{E}[g_{m+1}] \leq H(\mu_{m+1}, \delta_{m+1})$. Since $\mu_{m+1} = y^*/\gamma - y_0^*$, simple calculations (e.g., from [42]) show that in the range $y_0^* \in [0, y^*/(2\gamma))$, $H(\mu_{m+1}, \delta_{m+1})$ is maximized when $y_0^* = 0$. Thus, $\mathbf{E}[g_{m+1}] \leq e^{-y^*/(8\gamma)}$. Using this bound on $\mathbf{E}[g_{m+1}]$ along with (12), we see from Theorem 5.2 that it suffices to have

$$(1 - \min\{(B/\gamma)^2 \cdot (e^2 + \max\{2, (B-1)^{-1}\}), 1\})^m > e^{-y^*/(8\gamma)}. \tag{13}$$

Recall that $B \in (1,2)$. The recipe of Remark 7.1 shows that $\gamma = (K'B/(B-1)) \cdot \max\{m/y^*, 1\}$ for a sufficiently large constant $K' \geq 2$, satisfies (13). ∎

8.2 I-version: maximizing total weighted flow

We now present a more sophisticated use of the ideas of Section 7. The main new point here will be that randomized rounding itself may be insufficient, in contrast with the previously seen algorithms. We will have

to do a final "cleanup" phase which can reduce the objective function; we will ensure, however, that this reduction is not much.

We consider the I-version defined in Section 2; the setting is somewhat similar to that of Section 7. We are given a digraph $G = (V, E)$ and disjoint subsets S and C of V. Each edge $(u, v) \in E$ has a capacity $c_{u,v} > 0$; each $t \in C$ has a demand $d_t > 0$ and a weight $w_t > 0$. By scaling, we assume that the weights lie in $[0, 1]$. Let $|V| = n$ and $|E| = m$ as before. Informally, the main new constraint is that the flows constructed must be *integral*; the objective is to maximize the total weighted flow. More formally, recalling the discussions of Sections 7 and 7.1, our problem is the following ILP. Let $C = \{t_1, t_2, \ldots, t_k\}$. The ILP is to construct k non-negative flows $f^{(1)}, f^{(2)}, \ldots f^{(k)}$ on G such that:

(D1) in each $f^{(i)}$, flow-conservation is satisfied at all nodes in $V - (S \cup \{t_i\})$, and t_i is the only node that can have a positive in-flow;

(D2) the total flow on any edge (u, v) is at most $c_{u,v}$, i.e., $\sum_{i \in [k]} f^{(i)}_{u,v} \leq c_{u,v}$;

(D3) $\forall i,\ f^{(i)}_{in}(t_i) \leq d_i$; and

(D4) $\forall (i, u, v),\ f^{(i)}_{u,v}$ is a non-negative integer.

For each $i \in [k]$, define $x_i = f^{(i)}_{in}(t_i)/d_i$; note that $x_i \in [0, 1]$. The objective is to maximize $\sum_{i \in [k]} w_i x_i$. By (D4), $\sum_{i \in [k]} f^{(i)}_{u,v}$ must be an integer for all (u, v); thus, we may replace $c_{u,v}$ by $\lfloor c_{u,v} \rfloor$, without changing the problem. Similarly, we may replace d_i by $\lfloor d_i \rfloor$ for all i. So we assume henceforth that all the capacities and demands are positive integers. (Zero-capacity edges and zero-demand elements of C can be removed.)

LP relaxation and randomized rounding. The LP relaxation (LP2) that we consider, removes the constraints (D4) and lets the variables $f^{(i)}_{u,v}$ be non-negative reals. Letting y^* denote the LP optimum, we show how to round this LP to get an integral solution of value $\Omega(y^* \cdot \min\{y^*, m\}/m)$.

Let x^*_i denote the value of x_i in (LP2)'s computed optimal solution; thus, $y^* = \sum_i w_i x^*_i$. As in Section 7.1, we conduct flow-decomposition on the LP solution. For each $t_i \in C$, this yields flow-paths $P_{i,1}, P_{i,2}, \ldots, P_{i_{\ell_i}}$, where each $P_{i,j}$ originates from some $s_{i,j} \in S$ and carries a flow of value $z^*_{i,j} \geq 0$; we have $\sum_j z^*_{i,j} = d_i x^*_i$. Note also that for all $(u, v) \in E$, $\sum_{(i,j):\ (u,v) \in P_{i,j}} z^*_{i,j} \leq c_{u,v}$. Furthermore, from property (B0) of Section 7, we can ensure that $\ell_i \leq m$ for each i. Let $z'_{i,j} = z^*_{i,j} - \lfloor z^*_{i,j} \rfloor$. We further decompose the above flow into its "integral part" $\{\lfloor z^*_{i,j} \rfloor : i \in [k], j \in [\ell_i]\}$ and its "fractional part" $\{z'_{i,j} : i \in [k], j \in [\ell_i]\}$. We install the integral part, and update each edge (u, v)'s capacity to its residual capacity $c'_{u,v} = c_{u,v} - \sum_{(i,j):\ (u,v) \in P_{i,j}} \lfloor z^*_{i,j} \rfloor$. For each $t_i \in C$, let $d'_i = d_i - \sum_j \lfloor z^*_{i,j} \rfloor$. We now focus on rounding the fractional part. It is easy to check that all the $c'_{u,v}$ and d'_i are non-negative integers. For the rounding, we can ignore all i such that $d'_i = 0$, and all (u, v) such $c'_{u,v} = 0$; thus we assume henceforth that all the d'_i and $c'_{u,v}$ are positive integers.

Our randomized rounding is to choose a parameter $\gamma > 1$, and *independently* for all pairs (i, j), to round the flow on path $P_{i,j}$ to: (i) 1 with probability $z'_{i,j}/\gamma$, and to: (ii) 0 with probability $1 - z'_{i,j}/\gamma$. Let $Z_{i,j} \in \{0, 1\}$ denote the resultant random integer. We need one extra step; note from (D3) that we require $\sum_j Z_{i,j} \leq d'_i$ for each $t_i \in C$. For each i such that $\sum_j Z_{i,j} > d'_i$, we delete integral amounts of flow in an arbitrary way from the paths $\{P_{i,j} : j \in [\ell_i]\}$, so that we have $f^{(i)}_{in}(t_i) = d'_i$.

Analysis of the rounding. Define $y^*_0 = \sum_i w_i (\sum_j \lfloor z^*_{i,j} \rfloor)/d_i$, and $y^*_1 = \sum_i w_i (\sum_j z'_{i,j})/d_i$. Note that $y^* = y^*_0 + y^*_1$. The integral part of the flow that we install, contributes y^*_0 to the objective function; the rounding contributes $\sum_i w_i (\sum_j Z_{i,j})/d_i$ minus the (weighted) amount of flow deleted. We will show how

to make the rounding contribute a net amount of $\Omega(y_1^* \cdot \min\{y_1^*, m\}/m)$. A simple case analysis shows that $y_0^* + \Omega(y_1^* \cdot \min\{y_1^*, m\}/m) \geq \Omega(y^* \cdot \min\{y^*, m\}/m)$. Thus, it will suffice for us to make the rounding contribute $\Omega(y_1^* \cdot \min\{y_1^*, m\}/m)$; this is what we focus on from now on. Henceforth, we ignore the installed integral flow, and only focus on the rounding: by "objective function", we will thus only mean the part of the final objective function that is contributed by the rounding. Let a^+ denote $\max\{a, 0\}$. Define $\alpha = \sum_i \sum_j w_i Z_{i,j}/d_i$; for each $i \in [k]$, let $\beta_i = w_i((\sum_j Z_{i,j}) - d_i')^+/d_i$. It is easy to see that the objective function value is $\alpha - \sum_{i \in [k]} \beta_i$.

To lower-bound $\alpha - \sum_{i \in [k]} \beta_i$, we aim to suitably lower-bound α and upper-bound $\sum_{i \in [k]} \beta_i$. We also need to ensure that no edge's capacity is violated: since all the $c'_{u,v}$ and $Z_{i,j}$ are integers, edge (u, v)'s capacity is exceeded only if $\sum_{(i,j):\ (u,v) \in P_{i,j}} Z_{i,j} \geq c'_{u,v} + 1$. Thus, in order to show that the objective function value is at least, say, $y_1^*/(4\gamma)$, it will suffice to simultaneously avoid the following $m + 2$ "bad" events:
(a) Event $E_{u,v}$, for edge (u, v): "$\sum_{(i,j):\ (u,v) \in P_{i,j}} Z_{i,j} \geq c'_{u,v} + 1$". We have m such events.
(b) Event E': "$\sum_{i \in [k]} \beta_i \geq y_1^*/(4\gamma)$".
(c) Event E'': "$\alpha \leq y_1^*/(2\gamma)$".

The underlying independent binary random variables in the sense of Theorem 5.2, are now $\{Z_{i,j} : i \in [k], j \in [\ell_i]\}$. Also, the first $m + 1$ events above are increasing, while the last event E'' is decreasing. We will construct suitable well-behaved estimators for our $m + 2$ bad events. Note that α is a sum of independent r.v.s $w_i Z_{i,j}/d_i$, each lying in $[0, 1]$; also, $\mathbf{E}[\alpha] = y_1^*/\gamma$. Thus, as in the proof of Theorem 7.2, $g'' = 2^{y_1^*/(2\gamma) - \alpha} = 2^{y_1^*/(2\gamma)} \cdot \prod_{i,j} 2^{-w_i Z_{i,j}/d_i}$ is a well-behaved estimator for E''; bound (3) shows that

$$\mathbf{E}[g''] \leq e^{-y_1^*/(8\gamma)}. \tag{14}$$

The events $E_{u,v}$ are also simple to handle. Fix $(u, v) \in E$; we have $\mathbf{E}[\sum_{(i,j):\ (u,v) \in P_{i,j}} Z_{i,j}] \leq c'_{u,v}/\gamma$. So by (2), there is a well-behaved estimator

$$g_{u,v} \equiv [\gamma(1 + 1/c_{u,v})]^{-c'_{u,v} - 1} \cdot \prod_{(i,j):\ (u,v) \in P_{i,j}} [\gamma(1 + 1/c_{u,v})]^{Z_{i,j}}$$

for $E_{u,v}$, with $\mathbf{E}[g_{u,v}] \leq (\frac{ec'_{u,v}}{\gamma(c'_{u,v}+1)})^{c'_{u,v}+1}$. If we ensure that $\gamma \geq 2e$, then this bound is maximized when $c'_{u,v} = 1$ (recall that $c'_{u,v}$ is a positive integer). Thus,

$$\mathbf{E}[g_{u,v}] \leq (e/(2\gamma))^2, \text{ if } \gamma \geq 2e. \tag{15}$$

The event E' needs a little more work:

Lemma 8.2 *Suppose $\gamma \geq 2e$. Then there is an explicit well-behaved estimator g' for the event E', such that $\mathbf{E}[g'] \leq 2e^2/\gamma$.*

Proof. It is immediate that $(4\gamma/y_1^*) \cdot \mathbf{E}[\sum_{i \in [k]} \beta_i]$ is a well-behaved estimator for E'. We next express $\mathbf{E}[\sum_{i \in [k]} \beta_i]$ in a more tractable form. Fix any $i \in [k]$, and recall that the number ℓ_i of flow-paths $P_{i,j}$ is at most m. Thus, $\sum_j Z_{i,j}$ can be at most m. We have

$$\mathbf{E}[\beta_i] = \mathbf{E}[w_i((\sum_j Z_{i,j}) - d_i')^+/d_i] = \sum_{s \in [m]} \frac{w_i}{d_i} \cdot \Pr(\sum_j Z_{i,j} \geq d_i' + s); \tag{16}$$

this now leads us to a well-behaved estimator for $\mathbf{E}[\sum_{i \in [k]} \beta_i]$. Define $x_i' = (\sum_j z_{i,j}')/d_i'$, and note that $x_i' \in [0, 1]$. We have $\mathbf{E}[\sum_j Z_{i,j}] = d_i' x_i'/\gamma$. Consider any $s \geq 1$. By (2), there is a well-behaved estimator

$$g_{i,s}' \equiv [\gamma(d_i' + s)/(d_i' x_i')]^{-d_i' - s} \cdot \prod_{j \in [\ell_i]} [\gamma(d_i' + s)/(d_i' x_i')]^{Z_{i,j}}$$

for the event "$\sum_j Z_{i,j} \geq d'_i + s$", with

$$\mathbf{E}[g'_{i,s}] \leq (\frac{ex'_i d'_i}{\gamma(d'_i + s)})^{d'_i + s}.$$

We get

$$\sum_{s \in [m]} \mathbf{E}[g'_{i,s}] \leq \sum_{s \geq 1} (\frac{ex'_i d'_i}{\gamma(d'_i + 1)})^{d'_i + s} \leq 2(\frac{ex'_i d'_i}{\gamma(d'_i + 1)})^{d'_i + 1} \leq (ex'_i/\gamma)^2/2;$$

the last two inequalities follow from the fact that that $\gamma \geq 2e$. So, by (16), we have an explicit well-behaved estimator $g' \doteq (4\gamma/y_1^*) \cdot \sum_{i \in [k]} \sum_{s \in [m]} w_i g'_{i,s}/d_i$ for the event E', with

$$\mathbf{E}[g'] \leq (4\gamma/y_1^*) \cdot [\sum_{i \in [k]} \frac{w_i}{d_i} (ex'_i/\gamma)^2/2] \leq \frac{2e^2}{\gamma y_1^*} \cdot \sum_{i \in [k]} w_i (x'_i)^2. \tag{17}$$

Finally, to upper-bound the last term of (17), recall that $y_1^* = \sum_{i \in [k]} w_i x'_i$, with all the x'_i lying in $[0, 1]$. Thus, the last term in (17) is at most $2e^2/\gamma$. □

Now that we have (14), (15) and Lemma 8.2, Theorem 5.2 shows that it suffices for us to take $\gamma \geq 2e$ and ensure that

$$(1 - \min\{2e^2/\gamma, 1\}) \cdot (1 - (e/(2\gamma))^2)^m > e^{-y_1^*/(8\gamma)}$$

holds. Recalling Remark 7.1, we can check that these requirements hold if $\gamma = K \cdot \max\{m/y_1^*, 1\}$, where K is a sufficiently large absolute constant. This completes our analysis.

Theorem 8.3 *For the multi-source, multi-sink* I-version *where we wish to maximize the total weighted flow, we can deliver a solution of value* $\Omega(y^* \cdot \min\{y^*, m\}/m)$ *in polynomial time.*

Finally, a hardness result of [21] shows (even for the case where all the capacities, demands, and weights lie in $\{0, 1\}$) that for any fixed $\epsilon > 0$, approximating this problem to within $m^{1/2-\epsilon}$ on directed graphs is *NP*-hard. Thus, Theorem 8.3 is essentially best-possible.

9 Discussion and Concluding Remarks

How do the results presented here relate to the original policy questions we started investigating? The positive and negative results are a step toward providing policy-makers a quantitative and mathematical basis for establishing policies. In particular, many of the problems considered have not been studied in the power engineering community until recently. This is largely due to the fact these problems were previously not practically relevant due to the regulated structure of the industry. Our intractability results demonstrate the hardness of the problems for extremely simple networks. Our results also show that the underlying network as well as the spatial distribution of sources and sinks play an important role in contract satisfaction problems arising in deregulated environments. For instance, in the regime of regulated markets, the Unit Commitment and Economic Dispatch problems could be solved optimally by individual companies. In contrast, in a deregulated market, implementation of policies by the ISO seeking to effectively use the network capacity under contractual constraints might be computationally intractable. An example of this is finding the maximum number of satisfiable contracts: the equivalent combinatorial problem being (0/1-VERSION, MAX-#CONTRACTS). Our research also allowed us to formulate a number of variants of the problems originally suggested. Some of the variants appear to be interesting from a computational standpoint. We expect that some of these and further

variants will appeal to practitioners as reasonable models of the actual problems. We have shown that some of these variants have polynomial-time approximation algorithms with worst-case performance guarantees. In [13, 14], we have begun an experimental evaluation of the algorithms presented here on realistic power networks. The goal is to measure the effectiveness of heuristics for various contract scenarios on a realistic power network. To test the algorithms in a realistic setting, we are working with an aggregated version of the Colorado Power Network. The contract scenarios are generated to capture various intuitively plausible "what if" scenarios, such as particular producers providing extremely cheap power, sudden increase in demand, etc.

An obvious open question is to try to improve the performance bounds for the problems considered. The most important problem in this context is to settle the approximability of the MULTI-SOURCE-(0/1-VERSION, MAX-FEASIBLE FLOW) problem. To see why any direct extension of randomized rounding based techniques might not work, let us go back to Example 2 given in Section 3. As discussed there, the source and the destination of a contract might not lie in the same connected component of the network obtained by putting in all arcs carrying positive flow. This is true not only of the linear programming relaxation we have given, but even for an optimal solution strategy. New approaches to tackle these issues will be of much interest.

We have begun a multi-year program to develop a comprehensive and detailed simulation to study the effects of deregulation of the electrical power industry. The goal is to include the entire North American continent and perform the simulation at a very high level of fidelity; in particular to include *each* significant element of the transmission system including generators, transmission elements, varied control elements and load distribution buses. The planned system is a hierarchical multi-resolution simulation hierarchy. A distinguishing feature of this system is an associated market model that will represent the customers, system operators and individual companies. The market model is coupled to a physical model of the power grid. More details on the project will be available (see [12, 24] for a preliminary description on the subject).

Acknowledgments: We thank Dr. Martin Wildberger (EPRI) for introducing us to the problems considered in this paper. Much of the work reported here would not have been possible without his support and the extensive discussions with him on topic related to this paper. We also thank Massoud Amin (EPRI), Terence Kelly (University of Michigan), R. Ravi (CMU), Suvrajeet Sen (University of Arizona), and the members of the Power Simulation project at Los Alamos National Laboratory: Darrell Morgensen, Dale Henderson, Jonathan Dowell, Martin Drozda, Verne Loose, Achla Marathe Doug Roberts, Gerald Sheble (Iowa State University) and Chris Barrett for fruitful discussions on topics related to the subject of this paper. Finally, we thank the reviewer of an earlier version of this work for her/his detailed comments.

References

[1] S. Arora and M. Sudan. Improved low-degree testing and its applications. In *Proc. 29th Annual ACM Symposium on Theory of Computing*, pp. 485-496 (1997).

[2] G. Ausiello, P. Crescenzi, G. Gambosi, V. Kann, A. Marchetti-Spaccamela and M. Protasi. *Complexity and approximation: Combinatorial optimization problems and their approximability properties*, Springer Verlag, 1999.

[3] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin. *Network flows: theory, algorithms, and applications.* Prentice Hall, Englewood Cliffs, New Jersey, 1993.

[4] J. Bard. Short term scheduling of thermal-electric generators using Lagrangian relaxation. *Operations Research*, 36(5), 1988.

[5] D. Bertsekas, G. Lauer, N. Sandell and T. Posberg. Solution of large scale optimal unit commitment problems. *IEEE Transactions on Power Apparatus and Systems*, PAS 101(1), 1982.

[6] R. Bohn, M. Caramanis and F. Schweppe. Optimal Pricing in Electrical Networks Over Space and Time. *Rand J. on Economics*, 18(3), 1984.

[7] California Public Utilities Commission. Order Instituting Rulemaking on the Commission's Proposed Policies Governing Restructuring California's Electric Services Industry and Reforming Regulation, Decision 95-12-063, Dec. 1995.

[8] H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Annals of Mathematical Statistics*, 23:493–509, 1952.

[9] T. Cormen, C. Leiserson, and R. Rivest. *Introduction to Algorithms.* MIT Press, Cambridge, MA, 1990.

[10] P. Crescenzi and V. Kann. A compendium of *NP* optimization problems, Manuscript, KTH, Sweden, 1995.

[11] A. Cohen and V. Sherkat. Optimization Based Methods for Operation Scheduling. *Proceedings of the IEEE*, 75(12):1574-1591, 1987.

[12] C. Barrett, A. Marathe and M. Marathe. Parametric Models for Large Scale Simulation of Deregulated Electrical Power Industry. Los Alamos Technical Report, 2001.

[13] D. Cook, V. Faber, M. Marathe, A. Srinivasan and Y. J. Sussmann. Combinatorial Problems in Production and Transmission of Electric Power: Theory and Experimental Results. Technical Report LAUR-99-0048, 1999, Los Alamos National Laboratory, Los Alamos, NM, 87545.

[14] D. Cook, G. Hicks, V. Faber, M. Marathe, A. Srinivasan, Y. J. Sussmann and H. Thornquist. Experimental Analysis of Contract Satisfaction Problems Arising in Deregulated Power Industry, to appear in *Proc. of Workshop on Algorithmic Engineering*, Arhus, Denmark, Auhgust 2001.

[15] Y. Dinitz, N. Garg and M. Goemans. On the Single-Source Unsplittable Flow Problem. In *Proc. IEEE Symposium on Foundations of Computer Science*, pp. 290–299, 1998.

[16] M. E. Dyer and A. M. Frieze. Planar 3DM is *NP*-complete. J. Algorithms, 7:174–184, 1986.

[17] The Changing Structure of the Electric Power Industry: Selected Issues. DOE 0562(98), Energy Information Administration, US Department of Energy, Washington, D.C., 1998.

[18] *The Economist*, Asian edition. From circuits to packets. in *Survey on Telecommunications*, pp. 25–27, Sept. 13–19, 1997.

[19] EPRI-Workshop: Underlying Technical Issues in Electricity Deregulation. Technical report forthcoming, Electric Power Research Institute (EPRI), April 25-27, 1997.

[20] The U.S. Federal Energy Regulatory Commission. Notice of Proposed Rulemaking ("NOPRA"). Dockets No. RM95-8-000, RM94-7-001, RM95-9-000, RM96-11-000, April 24, 1996.

[21] V. Guruswami, S. Khanna, R. Rajaraman, B. Shepherd and M. Yannakakis. Near-optimal hardness results and approximation algorithms for edge-disjoint paths and related problems. In *Proc. ACM Symposium on Theory of Computing*, pages 19–28, 1999.

[22] M. Hassoun and P. Watta. Optimization of the unit commitment problem by a coupled gradient network and by a genetic algorithm. Technical Report, TR-103697, Electric Power Research Institute (EPRI), Palo Alto, California, 1994.

[23] J. Håstad. Testing of the long code and hardness for clique. In *Proc. ACM Symposium on the Theory of Computing*, pp. 11–19, 1996.

[24] L. J. Dowell, M. Drozda, D. B. Henderson, V. Loose, M. Marathe, and D. Roberts ELISIMS: Comprehensive Detailed Simulation of the Electric Power Industry. Technical Report, LA-UR-98-1739, Los Alamos National Laboratory, 1998.

[25] S. Herman. Delmar's standard text-book of electricity. Delmar Publishing Company, Albany, NY, 1995.

[26] J. Hewlett, D. Hale, T. Luoung and R. Eynon. An exploration of network modeling: The case of NEPOOL. *Issues in Midterm Analysis and Forecasting*, DOE/EIA-0607(98), Washington, D.C., pp. 11–28, July, 1998.

[27] W. Hoeffding. Probability inequalities for sums of bounded random variables. *American Statistical Association Journal*, 58:13–30, 1963.

[28] W. Hogan. Contract networks for electric power transmission. *J. Regulatory Economics*, pp. 211-242, 1992.

[29] V. Kann. Maximum bounded 3-dimensional matching is MAX SNP-complete. Inform. Process. Lett. 37, pp. 27-35, 1991.

[30] J. Kleinberg. *Approximation algorithms for disjoint paths problems.* Ph.D. Thesis, Department of EECS, MIT, 1996.

[31] J. Kleinberg. Single-source unsplittable flow. In *Proc. IEEE Symposium on Foundations of Computer Science*, pp. 68–77, 1996.

[32] S. G. Kolliopoulos and C. Stein. Improved approximation algorithms for unsplittable flow problems. In *Proc. IEEE Symposium on Foundations of Computer Science*, pp. 426–435, 1997.

[33] S. G. Kolliopoulos and C. Stein. Approximating disjoint-path problems using greedy algorithms and packing integer programs. In *Proc. MPS Conference on Integer Programming and Combinatorial Optimization*, pp. 153–168, 1998. Lecture Notes in Computer Science 1412, Springer-Verlag.

[34] G. Lawton. In Search of Real-Time Internet Services. In *IEEE Computer*, pp. 14–16, November 1997.

[35] J. Muckstadt and S. Koenig. An application of Lagrangean relaxation to scheduling in power generation systems. *Operations Research*, 25(3), pp. 387–403, 1977.

[36] C. Pang and H. Chen. Optimal short-term unit commitment. *IEEE Transactions on Power Apparatus and Systems*, 1976.

[37] R. Rajaraman, J. Sarlashkar and F. Alvarado. The effect of demand elasticity on security prices for the poolco and multi-lateral contract networks. *IEEE Transactions on Power Systems*, 12(3), pp. 1177-1184, August 1997.

[38] P. Raghavan. Probabilistic construction of deterministic algorithms: approximating packing integer programs. *Journal of Computer and System Sciences (JCSS)*, 37, pp. 130–143, 1988.

[39] P. Raghavan and C. D. Thompson. Randomized rounding: a technique for provably good algorithms and algorithmic proofs. *Combinatorica*, 7 pp. 365–374, 1987.

[40] R. Raz and S. Safra. A sub-constant error-probability low-degree test, and a sub-constant error-probability PCP characterization of *NP*. In *Proc. 29th Annual ACM Symposium on Theory of Computing*, pp. 475-484 (1997).

[41] F. Schweppe, M. Caramanis, R. Tabors and R. Bohn. *Spot Pricing of Electricity*. Kluwer Academic Publishers, Boston, MA 1988.

[42] A. Srinivasan. Improved approximations of packing and covering problems. In *Proc. ACM Symposium on the Theory of Computing*, pp. 268–276, 1995.

[43] A. Srinivasan. Improved approximations for edge disjoint paths, unsplittable flow, and related routing problems. In *Proc. IEEE Symposium on Foundations on Computer Science*, pp. 416–425, 1997.

[44] A. Srinivasan. Approximation Algorithms via Randomized Rounding: A Survey. *Lectures on Approximation and Randomized Algorithms* (M. Karoński and H. J. Prömel, eds.), Series in *Advanced Topics in Mathematics*, Polish Scientific Publishers PWN, Warsaw, pages 9–71, 1999.

[45] A. M. Wildberger. Issues associated with real time pricing. Unpublished Technical Report, Electric Power Research Institute (EPRI), 1997.

[46] A. M. Wildberger. Brief Overview of power systems operations, & Planning from the point of view of Mathematical Modeling, Simulation and Computation with Selected References to Books, EPRI Reports, Available Software & Web Sites. Unpublished Manuscript, Electric Power Research Institute (EPRI), 1997.

[47] Websites: http://www.magicnet.net/∼metzler/page2d.html and http://www.enron.com/choice/dereg.fset.html (also see references therein). http://www.eia.doe.gov/ (Energy Information Administration)

[48] The Wall Street Journal, August 04 2000, pp. A2; August 14 2000, pp. A6; and August 21 2000.

[49] W. Snyder, J. Rayburn, and H. Powell. Dynamic programming approach to unit commitment. *IEEE Transactions on Power Systems*, 1987.

[50] A. J. Wood and B. F. Wollenberg. *Power Generation, Operation and Control*. John Wiley and Sons, 1996.

[51] F. Zhuang and F. Galiana. Towards a more rigorous and practical unit commitment by Lagrangian relaxation. *IEEE Transactions on Power Systems*, 1988.

Appendix: Proof of Theorem 5.5

Let $S_1 = \{i \in [\ell] : \ r_i \geq 1/2\}$ and $S_2 = \{i \in [\ell] : \ r_i < 1/2\}$; define $Y_1 = \sum_{i \in S_1} X_i$ and $Y_2 = \sum_{i \in S_2} r_i X_i$. Also let $r'_i = 2r_i$ for each $i \in S_2$ and let $Y'_2 = \sum_{i \in S_2} r'_i X_i$. We claim that

$$g \doteq \lambda^{Y'_2 - 2} + Y_1 Y_2/\epsilon + \sum_{i,j \in S_1 :\ i<j} X_i X_j$$

is a well-behaved estimator as required, for the event $\mathcal{E} \equiv (Z \geq (1+\epsilon))$. Recalling Remark 5.3, it suffices to consider the special case of $u = 0$, in the definitions of (P1)–(P3).

To see (P1), note that $\mathbf{E}[g] = \lambda^{-2}(\prod_{i \in S_2} \mathbf{E}[\lambda^{r'_i X_i}]) + \epsilon^{-1}\mathbf{E}[Y_1]\mathbf{E}[Y_2] + \sum_{i,j \in S_1 :\ i<j} \mathbf{E}[X_i X_j]$, which is efficiently computable. Similarly, property (P3) is easy to check.

We now show (P2). Suppose $\mathcal{E}$ holds. Then, if $Y_1 = 0$, we must have $Y_2 \geq 1 + \epsilon$, and hence must have $Y_2 \geq 1$; similarly, if $Y_1 = 1$, the event $Y_2 \geq 1$ should hold. Thus,

$$\begin{aligned}
\Pr(\mathcal{E}) &\leq \Pr(Y_1 = 0) \cdot \Pr(Y_2 \geq 1) + \Pr(Y_1 = 1) \cdot \Pr(Y_2 \geq \epsilon) + \Pr(Y_1 \geq 2) \\
&\leq \Pr(Y_2 \geq 1) + \mathbf{E}[Y_1] \cdot \mathbf{E}[Y_2]/\epsilon + \Pr(Y_1 \geq 2) \quad \text{(Markov's inequality)} \\
&= \Pr(Y'_2 \geq 2) + \mathbf{E}[Y_1] \cdot \mathbf{E}[Y_2]/\epsilon + \Pr(Y_1 \geq 2) \\
&\leq \mathbf{E}[\lambda^{Y'_2 - 2}] + \mathbf{E}[Y_1] \cdot \mathbf{E}[Y_2]/\epsilon + \Pr(Y_1 \geq 2) \quad \text{(by (2))} \\
&\leq \mathbf{E}[\lambda^{Y'_2 - 2}] + \mathbf{E}[Y_1] \cdot \mathbf{E}[Y_2]/\epsilon + \sum_{i,j \in S_1 :\ i<j} \Pr(X_i = X_j = 1). \qquad (18)
\end{aligned}$$

This proves (P2).

Finally, let us see that (18) can be bounded by $(e^2 + \max\{2, \epsilon^{-1}\})/\lambda^2$. From (2), we have $\mathbf{E}[\lambda^{Y'_2 - 2}] \leq G(2/\lambda, \lambda - 1) \leq (e/\lambda)^2$. Also, since $\mathbf{E}[Y_1] = \sum_{i \in S_1} \Pr(X_i = 1)$, it is not hard to check that

$$\sum_{i,j \in S_1 :\ i<j} \Pr(X_i = 1) \cdot \Pr(X_j = 1) \leq (\mathbf{E}[Y_1])^2/2.$$

Thus, we need only show that

$$\mathbf{E}[Y_1] \cdot \mathbf{E}[Y_2]/\epsilon + (\mathbf{E}[Y_1])^2/2 \leq \max\{2, \epsilon^{-1}\}/\lambda^2, \qquad (19)$$

which we do now. Let $\mathbf{E}[Z] = a > 0$, $\mathbf{E}[Y_1] = x \geq 0$, and $\mathbf{E}[Y_2] = y \geq 0$. We clearly have the constraints $x/2 + y \leq a \leq x + y$. Note in particular that $y \leq a - x/2$. Thus, the left-hand side of (19) is at most the maximum value of the function $f(x) = \epsilon^{-1}x(a - x/2) + x^2/2$, subject to the constraint that $0 \leq x \leq 2a$. Note that $f'(x) = \epsilon^{-1}(a + x(\epsilon - 1))$. If $\epsilon \geq 1/2$, $f'(x) \geq 0$ for $0 \leq x \leq 2a$; thus, the maximum of f is attained when $x = 2a$. Hence, $f(x) \leq 2a^2$. Next, if $\epsilon < 1/2$, f attains its maximum when $x = x^* = a/(1-\epsilon)$. Thus, $f(x) \leq f(x^*) = a^2/(2\epsilon(1-\epsilon)) \leq \epsilon^{-1}a^2$ here. So, the l.h.s. of (19) is at most $a^2 \cdot \max\{2, \epsilon^{-1}\}$. Recalling that $a \leq 1/\lambda$ completes the proof. ∎