Experimental Analysis of Algorithms for Bilateral-Contract Clearing Mechanisms Arising in Deregulated Power Industry

CHRIS BARRETT* DOUG COOK† VANCE FABER‡ GREGORY HICKS§
ACHLA MARATHE* MADHAV MARATHE* ARAVIND SRINIVASAN¶
YORAM J. SUSSMANN|| HEIDI THORNQUIST**

Abstract: We consider the bilateral contract satisfaction problem arising from electrical power networks due to the proposed deregulation of the electric utility industry in the USA. Given a network G and a set (or multiset) of pairs of vertices (contracts) with associated demand functions that reflect the amount of flow which needs to be sent between the pairs of vertices, the goal is to find the maximum number of simultaneously satisfiable contracts. Four different algorithms for bilateral contract satisfaction problems are considered: (i) SMALLEST-FIRST HEURISTIC (ii) LARGEST-FIRST HEURISTIC (iii) RANDOM-ORDER HEURISTIC and (iv) ILP-RANDOMIZED ROUNDING: an integer linear programming based approximation algorithm with proven performance guarantee in restricted cases. The main focus of the paper is to study how the heuristics performed in fairly realistic settings. For this purpose we used an approximate electrical power network from Colorado.

From our preliminary analysis, it appears that although the first three heuristic algorithms do not have a worst-case guarantee, they outperform the theoretically better ILP-RANDOMIZED ROUNDING algorithm. We also tested the algorithm on four types of *scenarios* that are likely to occur in a deregulated marketplace. The empirical results show that the networks that are adequate in regulated marketplace might not be adequate for satisfying all the bilateral contracts in a deregulated industry. Our studies indicate that simple clearing mechanisms that are currently in use in many of the power markets are computationally fast and *near-optimal* in their use of network capacity; also, a linear programming upper bound on the quality of any solution, turns out to be a very good bound in practice.

*Basic and Applied Simulation Sciences, (D-2), and Computer and Computational Science (CCS-3), Los Alamos National Laboratory, Los Alamos. Email: {barrett, achla, marathe}@lanl.gov

†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.

§Department of Mathematics, North Carolina State University. Email: hicksgp@yahoo.com.

¶Bell Laboratories, Lucent Technologies, 600 Mountain Ave., Murray Hill, NJ 07974-0636. Email: srin@research.bell-labs.com. Part of this work was done while at the National University of Singapore.

||Email: yorams@atl.fundtech.com.

**Department of Computational Science, Rice University, Houston, TX. Email: heidi@caam.rice.edu.

1 Introduction

The U.S. electric utility industry is undergoing major structural changes in an effort to make it more competitive. The step toward deregulation of the electric industry was supposed to help customers avoid unusually high bills. Under deregulation, the utilities had to give up their monopolies, sell many of their plants and buy electricity from a state-sanctioned wholesale market. Since the wholesale market was big and transparent enough to offer low and competitive prices it was expected that deregulation would bring the price down significantly [WSJ, Wi97a, EPR97, Web, FED97, We98]. One major consequence of the deregulation will be to decouple the controllers of the network from the power producers, making it difficult to regulate the levels of power on the network; consumers as well as producers will eventually be able to negotiate prices to buy and sell electricity. See [Wi97a, Wi97b, Wi97a]. In practice, deregulation is complicated by the facts that all power companies will have to share the same power network in the short term, with the network's capacity being just about sufficient to meet the current demand. To overcome these problems, most U.S. states have set up an ISO (independent system operator): a non-profit governing body to arbitrate the use of the network. The basic questions facing ISOs are how to decide which contracts to deny (due to capacity constraints), and who is to bear the costs accrued when contracts are denied. Several criteria/policies have been proposed and/or are being legislated by the states as possible guidelines for the ISO to select a maximum-sized subset of contracts that can be cleared simultaneously [Wi97b]. These include: (a) Minimum Flow Denied: The ISO selects the subset of contracts that denies the *least* amount of proposed power flow. This proposal favors clearing bigger contracts first. (b) First-in First-out: The contract that comes first gets cleared first; this is the least discriminating to the contractors. (c) Maximum Consumers Served: This clears the smallest contracts first and favors the small buyers whose interests normally tend to go unheard.

There are three important issues in deciding policies that entail specific mechanisms for selecting a subset of contracts: (i) *fairness* of a given policy to producers and consumers; (ii) the *computational complexity* of implementing a policy, and (iii) how *sound* a given policy is from an economic standpoint. (For instance, does the policy promote the optimal clearing price, the optimal use of network resources, etc.)

Here we focus on evaluating the efficacy of a given policy with regard to its computational resource requirement and network resource utilization. It is intuitively clear that (i) the underlying network, (ii) its capacity and topology and (iii) the spatial locations of the bilateral contracts on the network, will play an important role in determining the efficacy of these policies. We do not discuss here the fairness aspects and extensive details of the economic aspects of these policies: these are important topics and subjects of a companion paper. The work reported here was done as part of two inter-related projects at Los Alamos. The first project is to develop a mathematical and computational theory of simulations based on Sequential Dynamical Systems (SDSs). The second aims at developing a comprehensive coupled simulation[1] of the *deregulated* electrical power industry [DD+98]. See `http://tsasa.lanl.gov` for additional details regarding these and other socio-technical simulation projects at Los Alamos. To achieve this goal, we carry out the first experimental analysis of several algorithms for simultaneously clearing a maximal number of bilateral contracts. The algorithms were chosen according to: (a) provable performance, (b) ability to serve as a proxy for some of the above-stated policies and (c) computational requirement.

The algorithms studied are as follows; see § 3 for their specification. The ILP-RANDOMIZED ROUNDING algorithm has a provable performance guarantee under certain conditions. The computational resource requirement was quite high, but the approach also provides us with an upper bound on any optimal solution and proved useful in comparing the performance of the algorithms. The LARGEST-FIRST HEURISTIC is a proxy for the *Minimum Flow Denied* policy. The SMALLEST-FIRST HEURISTIC serves as a proxy for the *Maximum Contracts Served* policy. The RANDOM-ORDER HEURISTIC orders the contract in a random order and tries to clear the contracts in the random order. This algorithm was chosen as a proxy for the *First-in*

[1] This coupled simulation has a market component and a physical network component.

First-out policy. Such a policy is probably the most natural clearing mechanism and is currently in place at many exchanges including California.

We used a coarse representation of the Colorado electrical power network (see § 5), to qualitatively compare the four algorithms discussed above in fairly realistic settings. The realistic networks differ from random networks and structured networks in the following ways: (i) Realistic networks typically have a very low average degree. In fact, in our case the average degree of the network is no more than 3. The same is true for the western states interconnect and the Texas interconnect. (ii) Realistic networks are not very uniform. One typically sees one or two large clusters (downtown and neighboring areas) and small clusters spread out throughout. (iii) For most empirical studies with random networks, the edge weights are chosen independently and uniformly at random from a given interval. However, realistic networks typically have very specific kinds of capacities since they were constructed with particular design goal.

Tests and Results: From our preliminary analysis, it appears that although the simple heuristic algorithms do not have worst-case guarantees, they outperform the theoretically better randomized rounding algorithm. We tested the algorithms on four carefully chosen scenarios. Each scenario was designed to test the algorithms and the resulting solutions in a deregulated setting. The empirical results show that networks that are capable of satisfying all demand in a regulated marketplace can often be inadequate for satisfying all (or even a acceptable fraction) of the bilateral contracts in a deregulated market. Our results also confirm intuitive observations: e.g., the number of contracts satisfied crucially depends on the scenario and the algorithm.

Significance: As far as we are aware, this is the first study to investigate the efficacy of various policies for contract satisfaction in a deregulated power industry. Since it was done in fairly realistic settings, the qualitative results obtained here have implications for policy makers, especially when we are witnessing power shortages in California. Our analysis shows that a sudden reduction in generating capacity (scenario 1, which can be caused by generators going off-line due to planned or unplanned maintenance schedules), unanticipated increase in demand (scenario 2), and inadequate transmission capacity (scenario 3), all have significant impact in limiting the number of contracts cleared. For a state like California, where all the three scenarios are occurring simultaneously, it is natural to expect unsatisfied customers. To compare the algorithms in a quantitative and (semi-)rigorous way, we employ statistical tools and experimental designs. Although many of the basic tools are standard in statistics, the use of formal statistical methods in experimental algorithmics for analyzing/comparing the performance of heuristics has not been investigated to the best of our knowledge. We believe that such statistical methods should be investigated further by the experimental algorithmics community for deriving more quantitative conclusions when theoretical proofs are hard or not very insightful. Our results can also be applied in other settings, such as bandwidth-trading on the Internet. See, e.g., [BW]. Finally, to our knowledge, previous researchers have not considered the effect of the underlying network on the problems; this is an important parameter especially in a free-market scenario. Appendix A illustrates this point further.

2 Problem Definitions

We briefly define the optimization problems studied here. We are given an undirected network (the power network) $G = (V, E)$ with capacities c_e for each edge e and a set of source-sink node pairs $(s_i, t_i), 1 \leq i \leq k$. 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 find the following notation convenient to describe the problems. Denote the set of nodes by N. The contracts are defined by a relation $R \subseteq (N \times N \times \Re \times \Re)$ 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 $cost(A) = \beta$. Corresponding to the power network, 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 $cost(A)$. By this construction, we can assume without loss of generality that each node can participate in exactly one contract. This fact will be used later. 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. Given a power network $G(V,E)$, 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 $B = supply(s) = demand(t) = \sum_{A \in R} flow(A)$.

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) try and find a subset of contracts that are feasible. Combining these two alternatives we define the following types of "relaxed feasible" subsets of R.

Definition 1 *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.*

Note that case (3) of Definition 1 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. Finally, note 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. Basic optimization problems related to relaxed feasible sets are:

Definition 2 *Given a graph $G(V,E)$ and a contract set R, the (*0/1-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 a* 0/1-contract satisfaction *feasible set of contracts. Is the related (*0/1-VERSION, MAX-#CONTRACTS*) problem, we aim 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.*

Though such electric flow problems have some similarities with those from other practical situations, there are many basic differences such as reliability, non-distinguishability between the power produced by different generators, short life-time (due to the lack of adequate storage devices), line effects, etc. [WW96]. The variants of flow problems related to power transmission studied here are intuitively harder than traditional multi-commodity flow problems, since we *cannot distinguish between* the flow "commodities" (power produced by different generators). As a result, current solution techniques used to solve single/multi-commodity flow problems are not directly applicable to the problems considered here; see the discussion in Appendix A.

3 Description and Discussion of Algorithms

We work on the (0/1-VERSION, MAX-#CONTRACTS) problem here. Let n and m respectively denote the number of vertices and edges in the network G. In [CF+97], it was shown that (0/1-VERSION, MAX-#CONTRACTS) is *NP*-hard; also, unless *NP* $\subseteq ZPP$, it cannot be approximated to within a factor of $m^{1/2-\epsilon}$ for any fixed $\epsilon > 0$, in polynomial time. Thus, we need to consider good heuristics/approximation algorithms.

First, there are three simple heuristics. The SMALLEST-FIRST HEURISTIC considers the contracts in non-decreasing order of their demands. When a contract is considered, we accept it if it can be feasibly added to the current set of chosen contracts, and reject it otherwise. The LARGEST-FIRST HEURISTIC is the same, except that the contracts are ordered in non-increasing order of demands. In the RANDOM-ORDER HEURISTIC, the contracts are considered in a random order. See Appendix B for worst-case examples for these heuristics.

We next discuss an approximation algorithm of [CF+97]. This has proven performance only when all source vertices s_i are the same; however, this algorithm extends naturally to the multi-source case which we work on. The basic ideas are as follows; see Appendix C for a comprehensive discussion. We create an ILP formulation for the problem, with $x_i \in \{0,1\}$ being the indicator variable for satisfying the demand from s_i to t_i or not; non-negative real variables $z_{i,e}$ indicate the flow due to the contract between (s_i, t_i) on edge e of the network. We relax the condition "$x_i \in \{0,1\}$" to "$x_i \in [0,1]$" and solve the resultant LP; let y^* be the LP's optimal objective function value. The randomized rounding of the LP solution is as follows. For a large enough constant c, define $\lambda > 1$ to be

$$c/\sqrt{\epsilon}, \text{ if } y^* > m; \; cm/(y^*\epsilon), \text{ if } y^* \leq m \text{ and } \epsilon \in (0,1/2]; \; c(m/y^*)^{(1-\epsilon)/\epsilon}, \text{ if } y^* \leq m \text{ and } \epsilon \in (0,1). \quad (1)$$

We perform the following rounding steps. (a) Independently for each i, set a random variable Y_i to 1 with probability x_i/λ, and $Y_i := 0$ with probability $1 - x_i/\lambda$. (b) If $Y_i = 1$, we will choose to satisfy $(1-\epsilon)$ of (s_i, t_i)'s contract: for all $e \in E$, set $z_{i,e} := z_{i,e}(1-\epsilon)/x_i$. (c) If $Y_i = 0$, we choose to have no flow for (s_i, t_i): i.e., we will reset all the $z_{i,e}$ to 0.

A deterministic version of this result based on *pessimistic estimators*, is also provided in [CF+97].

Theorem 3.1 ([CF+97]) *Given a network G and a contract set R, we can find an approximation algorithm for* (0/1-VERSION, MAX-#CONTRACTS) *when all source vertices are the same, as follows. Let OPT be the optimum value of the problem, and m be the number of edges in G. Then, for any given $\epsilon > 0$, we can in polynomial time find a subset of contracts R' with total weight $\Omega(OPT \cdot \min\{(OPT/m)^{(1-\epsilon)/\epsilon}, 1\})$ such that for all $i \in R'$, the flow is at least $(1-\epsilon)d_i$.*

4 Implementation details

We now discuss the key implementation details. For the three greedy heuristics the implementations are fairly straightforward, and we used public-domain network flow codes. Implementing the randomized rounding procedure requires extra care. The pessimistic estimator approach of [CF+97] works with very low probabilities, and requires significant, repeated re-scaling in practice. Thus we focus on the randomized version of the algorithm of [CF+97]. Consider λ and ϵ: an important empirical question is the optimal choice of these parameters. Equation (1) gives evidence of an inverse relationship between ϵ and λ. The key clue to specifying values of the parameter λ is provided by analyzing its purpose once more. The algorithm sets $Y(i) = 1$ with probability $x(i)/\lambda$ and $Y(i) = 0$ with probability $1 - x(i)/\lambda$. In other words, λ is used to weigh the probability in favor of $Y(i) = 0$, which drops off certain flows. Clearly, this increases the probability of satisfying the LP constraints. However, this also decreases the value of the objective function $OBJ = \sum_i Y(i)$ that we are attempting to maximize. Thus, the problem is to find a balance for λ.

Note that since $E[OBJ] = E[\sum_i Y(i)] = y^*/\lambda$, by applying Chernoff-Hoeffding bounds, we get $Pr(OBJ \leq y^*/2\lambda) \leq e^{-y^*/8\lambda}$. So, $1 - Pr(OBJ > y^*/2\lambda) \leq e^{-y^*/8\lambda}$; i.e., $Pr(OBJ > y^*/2\lambda) > 1 - e^{-y^*/8\lambda}$. Thus, if y^* is "large" and/or λ is small, the larger is the probability that our solution will be better than $1/(2\lambda)$ of the best possible solution. Hence, for say $\lambda = 1$ and $y^* \geq 104$ we have $e^{-y^*/8} \leq e^{-13}$ implying $Pr(OBJ > y^*/2) \geq 1 - e^{-13}$, which is approximately 1. That is, it is almost certain that the randomized rounding solution will be better than half the optimal relaxed solution. This implies that the randomized rounding solution may very well be a fairly good estimate of the optimal solution to the IP. However, it may certainly take an intractable amount of time to obtain such a feasible solution, as the probability of violating a constraint could now be exceedingly high and a suitable bound may not be able to be placed on them. This might seem natural to assume since the probability of $Y_i = 1$ would be x_i and $E[\sum_i Y_i] = y^* \geq OPT(IP)$ implying frequent infeasibilities could occur. So $\lambda = 1$ could in some cases be to too small and in others (larger ϵ) it may provide quality solutions. It can be seen that λ, say greater than 100, would be a poor choice since $Pr(Y_i = 0)$ is necessarily larger than 0.99 and thus $Pr(OBJ = 0) = Pr(\bigwedge_i(Y_i = 0)) = \Pi_i(Pr(Y_i = 0)) \geq (0.99)^k$ (where k is the number of initial contracts generated). So $Pr(OBJ = 0) \geq 0.22$ at $k = 150$ (a size in the range usually observed by the scenario generators) contracts, which is clearly not very promising. With this in mind, it makes sense to start at some $\lambda \leq 100$ and decrease the parameter's value until the runs on scenario instances show unsatisfying constraint violation. From this the above mentioned middle ground for λ may be reached. This inspires the following question. What is an unsatisfying constraint violation? Is it acceptable for the randomized rounding procedure to output a solution 1 out of every 5 times? For smaller λ it should be harder to obtain a feasible solution, but OBJ will probably be better. So the issue is that of running time. So, what is a reasonable proportion of feasibility success? This question needs to be considered for successfully performing the experiments. For this study, values of λ for which the proportion of success was 1 out of 5 was sought.

We also need to consider the role of ϵ in our experiments. A feasible solution in our context consists of sinks whose demand is satisfied to within an $(1 - \epsilon)$ fraction of their total request. We believe that $\epsilon > 1/2$ is not useful since it does not seem justifiable to satisfy only 50% of a county's household power demand. *We assume that for valid operation ϵ should typically be very close to 0.* For the sake of experimental analysis to determine an optimal λ, ϵ is fixed and a range of λ values is explored. This process is performed on a range of $\epsilon < 1/2$, taking more values of ϵ nearer to 0. This is easily accomplished by incrementing ϵ from .1 to .5 by some constant. The search for a λ (for each fixed ϵ) that realizes near-optimal performance of the randomized rounding procedure was automated in the implementation used for this analysis via bisection search (see PROCEDURE OPTIMIZE-λ below).

Notice that since the interval of bisection is fixed, the number of iterations the search algorithm is deterministic being the smallest integral solution to the inequality $k/2^{(x-1)} - k/2^x \leq \delta - \lambda$. So, for this study 12 iterations are made for each problem instance. For each problem instance an ordered pair (ϵ, λ) is created. Preliminary evaluations indicate that there is correlation between ϵ and λ and hence this data could possibly be used in conjunction with regression techniques to produce a relation equation which could perhaps, given an ϵ and a scenario type, predict with some accuracy an appropriate λ to use in the randomized rounding procedure. This possibility was not explored for this study, only five ϵ's were used per scenario and the last valid iteration of the bisection process was taken as the randomized rounding run.

PROCEDURE OPTIMIZE-λ :

1. For a given ϵ, set a value k for λ that is large enough to ensure feasibility for any generated problem instance and denote it λ_{top}. For our scenarios, we found $\lambda = 3$ was sufficiently large.
2. Let $\lambda_{bot} = 0$.
3. Run the randomized rounding procedure with λ_{top}.
4. **If** the run produces feasible solution, **then** $\lambda_{old} = \lambda_{top}$ and $\lambda_{top} = (\lambda_{top} + \lambda_{bot})/2$ **else** $\lambda_{bot} = \lambda_{top}$ and $\lambda_{top} = (\lambda_{old} + \lambda_{top})/2$.
5. If $(\lambda_{top} - \lambda_{bot}) \leq \delta - \lambda$, where $\delta - \lambda$ is a threshold set by the user, stop. Otherwise, go to Step 3. For this study, $\delta - \lambda = 0.001$.

5 Experimental Setup and Methodology

To test our algorithms experimentally, we used a network corresponding to a subset of a real power network along with contracts that we generated using different scenarios. The network we used is based on the power grid in Colorado and was derived from data obtained from PSCo's (Public Service Company of Colorado) Draft Integrated Resources Plan listing of power stations and major sub stations. The network is shown in Figure 2 (See Appendix). We restricted our attention to major trunks only. The plant capacities were derived from the PSCo data. To generate maximum aggregate demands for the sink nodes we used data from the US Census bureau on the number of household per county. The demands were generated by assigning counties to specific sink nodes in the network.

All the test cases were generated from the basic model. The general approach we used was to fix the edge capacities and generate source-sink contracts combinations using the capacities and aggregate demands in the basic model as upper bounds. To ensure that the test cases we generated corresponded to (1) difficult problems, i.e. infeasible sets of contracts, and (2) problems that might reasonably arise in reality, we developed several scenarios that included an element of randomness.

The current implementation is still based upon a network which should be feasible only if the total source capacity is greater than the total sink capacity and the only requirement is that the total sink capacity be satisfied regardless of which source provides the power. **Scenarios 1, 2, 3** and **4** are based around the network with total generating capacity, **6249 MW**, and reduced sink capacities near **4400MW** combined.

Scenario 1: The source capacity of the network was reduced until the running the MAXFLOW code indicated that the maximum flow in the network to be slightly less than the demand.

Scenario 2: For this scenario, we took the basic network and increased the sink capacity while the source capacity remained fixed.

Scenario 3: The edge capacities were adjusted, reduced in most cases, to limit the network to a maximum flow of slightly more than 4400MW given its source and sink distribution. The network is feasible at the 4400MW level if the flow provided to supply the sinks is permitted to adjust as necessary within the capacity of the individual sources. This network with bilateral contracts is not feasible.

Scenario 4: For this scenario, we took the network of Scenario 3 and biased the selection of source nodes towards the lower valued source units.

Methodology: We work with the four scenarios and five 5 representative values of ϵ varying from .1 to .5. We believe that satisfying a contract partially so that a contract is assigned less than .5 of the required demand is not very realistic. For each scenario, and for each of the 5 values of ϵ, the programs implementing the algorithms under inspection produced 30 files from which the following information could be extracted: (i) the running times and solutions of all four algorithms; and (ii) the LP upper bound on the IP. The number 30 was chosen to ensure that a statistically "large" sample of each measure would be provided in order to make valid statistical inference. More attention is given to the quality-of-solution measure of an algorithm (rather than the running-time measure), since from a social standpoint, contract satisfaction may leave little room for finding solutions that are far from optimal. For a given algorithm $\mathcal{A}$, let $Value_{\mathcal{A}}$ denote the number of contracts that can be satisfied by $\mathcal{A}$. $p_{\mathcal{A}} = \frac{Value_{\mathcal{A}}}{\lfloor y^* \rfloor}$ provides a measure of the quality of the algorithm's solution. The value $\lfloor y^* \rfloor$ provides an upper bound on the objective function value. The objective of our experiments was to find out which, if any, of the algorithms discussed here performs better than the others, in terms of quality of solution and running time for different contract scenarios.

6 Results and Analysis

The performance of the algorithms w.r.t. running time and quality of solutions is summarized in Tables 1 and 2. Table 3 summarizes the results of the statistical tests. Due to lack of space, additional results can be found in the Appendix D.

6.1 General Conclusions

1. Although there exist instances where the three heuristics produce solutions as large as $\Omega(n)$ times the optimal fractional solution, most of our tests show that we could find integral solutions fairly close to optimal.

2. From our analysis we can conclude that in realistic situations, the randomized rounding algorithm (basic version, without pessimistic estimators) is not likely to perform as well as the simple heuristics, both in terms of the quality of solution and running time.

3. Our experiments show that different scenarios make a *significant* difference in the type of solutions obtained. For example, the quality of solution obtained using the fourth scenario is significantly worse than the first three scenarios. The sensitivity to the scenarios poses interesting questions for infrastructure investment. The market will have to decide the cost that needs to be paid for expecting the necessary quality of service. It also brings forth the equity-benefit question: namely, who should pay for the infrastructure improvements?

4. It is possible that for certain scenarios, the underlying network is incapable of supporting even an acceptable fraction of the bilateral contracts. This observation although fairly intuitive provides an extremely important message, namely, networks that were adequate to service customers in a completely regulated power market *might* not be adequate in deregulated markets. This makes the question of evicting the bilateral contracts more important.

5. Intuitively, one would expect a trade-off between the number of contracts satisfied and the ϵ value. As ϵ increases, and the demand condition is more relaxed, a higher number of contracts should get satisfied. But our experiments show that the change in the number of contracts satisfied for different values of ϵ is insignificant.

6. The LP upper bound is a very useful tight bound in our results: some of our heuristics produce solutions very close to it.

7. In practical situations, the Random-Order heuristic would be the best to use since it performs very close to the optimal in terms of quality of solution and has very low running time.

8. Finally, in our analysis, $\lambda = 2$ gives the best solutions.

6.2 Statistical Analysis

We use a statistical technique known as *analysis of variance* (ANOVA) to test whether differences in the sample means of algorithms and scenarios reflect differences in the means of the statistical populations that they came from or are just sampling fluctuations. This will help us identify which algorithm and scenarios perform the best. See ([Fr68], [GH96]) for details.

Mathematical Model Quality of Solution: We first describe the experiment for the quality of solution, i.e., $p_{\mathcal{A}}$. We use a two-factor ANOVA model since our experiment involves two factors: the algorithms $\mathcal{A}_i$, $i = 1, 2, 3$ and 4, and scenarios $\mathcal{S}_j$, $j = 1, 2, 3$ and 4. Following classical statistics terminology, we will sometimes refer to algorithms as *treatments* and the scenarios as *blocks*. We will use $\mathcal{A}$ to denote the set of algorithms and $\mathcal{S}$ to denote the set of scenarios. For each algorithm-scenario pair we have *30 observations (or replicates)*. When testing the efficacy of the algorithms, we use 4 algorithms, each having 120 observations (30 for each scenario) from the corresponding population. The design of experiment used here is a *fixed-effect complete randomized block*. *Fixed-effect* because the factors are fixed as opposed to randomly drawn from a class of algorithms or scenarios. The conclusions drawn from this model will hold only for these particular algorithms and scenarios. *Complete* implies that the number of observations are the same for each block. *Randomized* refers to the 30 replicates being drawn randomly. We wish to test the hypothesis:

> Is the mean quality of solution provided by different algorithms the same, against the alternative hypothesis that some or all of these means are unequal?

The model for randomized block design includes constants for measuring the scenario effect (block effect), the algorithm effect (treatment effect) and a possible interaction between the scenarios and the algorithms. An appropriate mathematical model for testing the above hypothesis is given by: $X_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk}$. where X_{ijk} is the measurement ($p_{\mathcal{A}}$) for the kth sample within the ith algorithm and the jth scenario. τ_i is the algorithm effect. β_j is the scenario effect. $(\tau\beta)_{ij}$ captures the interaction present between the algorithms and the scenarios. ε_{ijk} is the random error. We use S-Plus [SPL98] software to run two-factor ANOVA to test the following three different hypothesis.

1. Are the means given by the 4 different algorithms equal i.e. $H_0 : \tau_1 = \tau_2 = \tau_3 = \tau_4$.

2. Are the means given by the 4 different scenarios equal i.e. $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4$.

3. Is there any interaction between the two factors i.e. $H_0 : (\tau\beta)_{ij} = 0$.

The results of two-factor ANOVA are shown in Table 3. In the following discussion, we explain the meaning of each column. DF refers to the degrees of freedom, SS refers to the sum of squared deviations from the mean. MS refers to the mean square error, which is the sum of squares divided by the degrees of freedom. The sum of squares for the algorithm factor can be calculated as: $SS_{\mathcal{A}} = nJ\Sigma_i(\overline{X}_{i\cdot\cdot} - \overline{X}_{\cdots})^2$ where n is the number of replicates, J is the number of scenarios, $\overline{X}_{i\cdot\cdot}$ is the mean of algorithm i across all scenarios and $\overline{X}_{\cdots}$ is the grand mean across all algorithms and scenarios. Recall that in our case $n = 30$ and $J = 4$ yielding a total sample size of 120.

The sum of squares for scenario factor can be calculated as: $SS_{\mathcal{S}} = nI\Sigma_j(\overline{X}_{\cdot j\cdot} - \overline{X}_{\cdots})^2$ where as before n is the number of replicates, I is the number of algorithms and $\overline{X}_{\cdot j\cdot}$ is the mean of scenario j across all algorithms. Again, in our case $n = 30$ and $I = 4$. The sum of squares for algorithms and scenario interaction is: $SS_{\mathcal{AS}} = n\Sigma j\Sigma_i[\overline{X}_{ij\cdot} - (\overline{X}_{\cdots} + \hat{\tau}_i + \hat{\beta}_j)]^2$. Here $\overline{X}_{ij\cdot}$ is the mean of observations for the algorithm i scenario j pair and $\hat{\tau}_i$ and $\hat{\beta}_j$ are respectively the estimated least square values of τ_i and β_j. The sum of squares "within" refers to the squared difference between each observation and the mean of the scenario and algorithm of which it is a member. It is also referred as the residual sum of squares. This can be calculated as: $SS_{\mathcal{W}} = n\Sigma j\Sigma_i\Sigma_k(X_{ijk} - \overline{X}_{ij\cdot})^2$. The total sum of squares is $SS_{\mathcal{T}} = SS_{\mathcal{A}} + SS_{\mathcal{S}} + SS_{\mathcal{AS}} + SS_{\mathcal{W}}$. The

Performance Measure: Quality of Solution (in %)					
	RR	SF	LF	RO	Scenario Means
Scenario 1	$\overline{X}_{11\cdot}$=48.68	$\overline{X}_{21\cdot}$=99.73	$\overline{X}_{31\cdot}$=97.97	$\overline{X}_{41\cdot}$=97.78	$\overline{X}_{\cdot 1\cdot}$=86.02
Scenario 2	$\overline{X}_{12\cdot}$=46.91	$\overline{X}_{22\cdot}$=99.56	$\overline{X}_{32\cdot}$=98.38	$\overline{X}_{42\cdot}$=98.93	$\overline{X}_{\cdot 2\cdot}$=85.94
Scenario 3	$\overline{X}_{13\cdot}$=45.69	$\overline{X}_{23\cdot}$=99.25	$\overline{X}_{33\cdot}$=97.10	$\overline{X}_{43\cdot}$=98.82	$\overline{X}_{\cdot 3\cdot}$=85.22
Scenario 4	$\overline{X}_{14\cdot}$=46.99	$\overline{X}_{24\cdot}$=98.03	$\overline{X}_{34\cdot}$=88.65	$\overline{X}_{44\cdot}$=93.41	$\overline{X}_{\cdot 4\cdot}$=81.77
Algo. Means	$\overline{X}_{1\cdot\cdot}$=47.07	$\overline{X}_{2\cdot\cdot}$=99.14	$\overline{X}_{3\cdot\cdot}$=95.51	$\overline{X}_{4\cdot\cdot}$=97.24	$\overline{X}_{\cdots}$ = 84.74

Table 1: The Mean Values of the Quality of Solution.

Performance Measure: Running Time (in Megaticks)					
	RR	SF	LF	RO	Scenario Means
Scenario 1	$\overline{X}_{11\cdot}$=163.33	$\overline{X}_{21\cdot}$=41.23	$\overline{X}_{31\cdot}$=24.57	$\overline{X}_{41\cdot}$=25.50	$\overline{X}_{\cdot 1\cdot}$=63.66
Scenario 2	$\overline{X}_{12\cdot}$=218.23	$\overline{X}_{22\cdot}$=49.63	$\overline{X}_{32\cdot}$=29.73	$\overline{X}_{42\cdot}$=30.23	$\overline{X}_{\cdot 2\cdot}$=81.96
Scenario 3	$\overline{X}_{13\cdot}$=181.70	$\overline{X}_{23\cdot}$=45.70	$\overline{X}_{33\cdot}$=23.30	$\overline{X}_{43\cdot}$=26.43	$\overline{X}_{\cdot 3\cdot}$=69.28
Scenario 4	$\overline{X}_{14\cdot}$=184.33	$\overline{X}_{24\cdot}$=44.53	$\overline{X}_{34\cdot}$=27.00	$\overline{X}_{44\cdot}$=27.27	$\overline{X}_{\cdot 4\cdot}$=70.78
Algo. Means	$\overline{X}_{1\cdot\cdot}$=186.90	$\overline{X}_{2\cdot\cdot}$=45.27	$\overline{X}_{3\cdot\cdot}$=26.15	$\overline{X}_{4\cdot\cdot}$=27.36	$\overline{X}_{\cdots}$ = 71.42

Table 2: The Mean Values of the Running Time.

p-value gives the smallest level of significance at which the null hypothesis can be rejected.[2] The lower the p-value, the lesser the agreement between the data and the null hypothesis. Finally the F**-test** is as follows. To test the null hypothesis, i.e. the population means are equal, ANOVA compares two estimates of σ^2. The first estimate is based on the variability of each population mean around the grand mean. The second is based on the variability of the observations in each population around the mean of that population. If the null hypothesis is true, the two estimates of σ^2 should be essentially the same. Otherwise, if the populations have different means, the variability of the population mean around the grand mean will be much higher than the variability within the population. The null hypothesis in the F-test will be accepted if the two estimates of σ^2 are almost equal.

In a two-factor fixed-effect ANOVA, three separate F**-tests** are performed. Two tests for the factors and third for the interaction term. If F-ratio is close to 1, the null hypothesis is true. If it is considerably larger implying that the variance between means is larger than the variance within a population, the null hypothesis is rejected. The F distribution table should be checked to see if the F ratio is significantly large. The results in Table 3 show that all the above three null hypothesis are rejected at any significance level. This implies that the performance (measured by $p_{\mathcal{A}}$) of at least one of the algorithms is significantly different from the other algorithms. Also, different scenarios make a difference in the performance. Finally, the scenarios and the algorithms interact in a significant way. The interaction implies that the performance of the algorithms are different for different scenarios.

Refer to Appendix D for more details on what caused the rejection of null hypotheses and for a similar discussion on running time performance measure. All the above analysis was performed while keeping the value of ϵ constant at 0.1. The performance of the randomized rounding algorithm[3] does not change in any significant way, both in terms of $p_{\mathcal{A}}$ and running time when ϵ varied from 0.1 to 0.5. So all the above results hold for $\epsilon = 0.1, 0.2, 0.3, 0.4$ and 0.5; the results are available from the authors upon request.

[2]To obtain a p-value for say $F_{\mathcal{A}}$, the algorithm effect, we would look across the row associated with 3 degree of freedom in the numerator and 464 degrees of freedom in the denominator in the F-distribution table and find the largest value that is still less than the one obtained experimentally. From this value, we obtain a p-value of 0 for $F_{\mathcal{A}}$.

[3]The other heuristics do not depend on the value of ϵ.

Source	DF	SS	MS	F**-test**	p**-value**
Scenario (Block)	3	0.14	0.05	43.38	0
Algorithm (Treatment)	3	22.78	7.59	6792.60	0
Scenario:Algorithm	9	0.12	0.01	15.90	0
Residuals	464	0.40	.0008		
Total	479	23.45			

Table 3: **Results of Two-Factor ANOVA:** This table shows results of two-factor ANOVA where the factors are algorithms and scenarios. The measurement is the quality of solution, given by $p_{\mathcal{A}}$. The p-values show that the algorithm effect, scenario effect and the interaction between the algorithms and scenarios are all significant at any level of confidence.

Summary: The SF heuristic clears the most contracts, almost as good as the optimal but takes more running time as compared to LF and RO. However, it takes only a quarter of the time as compared to RR. As far as scenarios go, first scenario clears most contracts in the least amount of time. From a practical standpoint, the RO heuristic seems to be the best: it performs very well both in terms of running time and quality of solution, and is trivial to implement. It performs very close to optimal in terms of clearing contracts and yet takes minimal time to do it as compared to the other algorithms. The RR algorithm gives good theoretical lower bounds, but does not seem very appropriate for real-life situations where both time and a high fraction of contract satisfaction are important.

7 Discussion and Concluding Remarks

We carried out an empirical study to evaluate the quality and running time performance of four different market clearing mechanisms. One heuristic was based on using a relaxation of integer linear program followed by randomized rounding of the fractional solution to yield an approximate integral solution. Experiments suggest that a basic version of the algorithm is not likely to be practically useful, given the running time and the quality of solution produced. The result is not entirely unexpected; it has been observed that some approximation algorithms that are designed to work in the worst case may not have a very good average case behavior. We also studied three different simple heuristics: experimental results suggest that each is likely to perform better than the theoretically provable approximation algorithm. This is in spite of the fact that it is very easy to construct instances where the heuristics have unboundedly poor performance guarantee. One of the heuristics: the *random-order heuristic* was studied to emulate a simple first-come first-serve type clearing mechanism that is currently employed by the California ISO. The heuristic performs surprisingly well when compared to a bound on any optimal solution obtained via linear programming. The results suggest that this simple clearing mechanism currently employed might result in near-optimal utilization of the network resources. Our overall assessment is that for the purposes of developing large scale microscopic simulations of the deregulated power industry, the three heuristic methods give sufficiently good performance in terms of the quality of solution and the computational time requirement.

References

[AMO93] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin. *Network flows: theory, algorithms, and applications*. Prentice Hall, Englewood Cliffs, New Jersey, 1993.

[ASE92] N. Alon, J. H. Spencer, and P. Erdős. *The Probabilistic Method*. Wiley–Interscience Series, John Wiley & Sons, Inc., New York,, 1992.

[BW] Example sites for information on bandwidth trading:
`http://www.lightrade.com/dowjones2.html` and
`http://www.williamscommunications.com/prodserv/network/webcasts/bandwidth.html`

[BMR00] C. Barrett, H. Mortveit and C. Reidys. Elements of a theory of computer simulation III: equivalence of SDS. to appear in *Applied Mathematics and Computation*, 2001.

[BH+00] "Dichotomy Results for Sequential Dynamical Systems," C. Barrett, H. Hunt III, M. Marathe, S. Ravi, D. Rosenkrantz and R. Stearns, Technical Report No. LA-UR-00-5984, Los Alamos National Laboratory, submitted, 2001.

[Ba88] J. Bard. Short term scheduling of thermal-electric generators using Lagrangean relaxation. *Operations Research*, 36(5), 1988.

[BL+82] D. Bertsekas, G. Lauer, N. Sandell and T.A. Posberg. Solution of large scale optimal unit commitment problems. *Transactions on Power Apparatus and Systems*, PAS 101(1), 1982.

[BCS84] R. Bohn, M. Caramanis and F. Schweppe. Optimal Pricing in Electrical Networks Over Space and Time. *Rand J. on Economics*, 18(3), 1984.

[CHH97] Cardell J.B., Hitt C.C. and Hogan W.W. (1997), Market Power and Strategic Interaction in Electricity Networks. `Resource and Energy Economics`, vol 19, 109-137.

[Ca95] 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.

[CF+97] D. Cook, V. Faber, M. Marathe, A. Srinivasan and Y. J. Sussmann. Low Bandwidth Routing and Electrical Power Networks. *Proc. 25th International Colloquium on Automata, Languages and Programming (ICALP)*, Aalborg, Denmark, LNCS 1443, Springer Verlag, pp. 604-615, 1998.

[CS87] A. Cohen and V. R. Sherkat. Optimization Based Methods for Operation Scheduling. *Proceedings of the IEEE*, 75(12):1574-1591, 1987.

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

[DOE98] The Changing Structure of the Electric Power Industry: Selected Issues. DOE 0562(98), Energy Information Administration, US Department of Energy, Washington, D.C. 1998.

[Eco97] *The Economist*, Asian edition. From circuits to packets. in *Survey on Telecommunications*, pp. 25–27, Sept. 13–19, 1997.

[Fr68] H.C. Fryer 1968. *Concepts and Methods of Experimental Statistics*, Allyn and Bacon Inc.

[GH96] Glass G.V. and K.D.Hopkins 1996. *Statistical Methods in Education and Psychology*, third edition.

[GS72] Glass G.V., Peckham P.D. and J.R. Sanders. 1972. "Consequences of failure to meet assumptions underlying the fixed-effects analysis of variance and covariance", *Review of Educational Research*, 42, 237-88.

[EPR97] EPRI-Workshop: Underlying Technical Issues in Electricity Deregulation. Technical report forthcoming, Electric Power Research Institute (EPRI), April 25-27, 1997.

[FED97] 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.

[GJ79] M. R. Garey and D. S. Johnson. *Computers and Intractability. A Guide to the Theory of NP-Completeness*. Freeman, San Francisco CA, 1979.

[He95] S. Herman. Delmar's standard text-book of electricity. Delmar Publishing Company, Albany, NY, 1995.

[HH98] 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.

[Ho92] W. Hogan. Contract networks for electric power transmission. *J. Regulatory Economics,* pp. 211-242, 1992.

[La97] G. Lawton, In Search of Real-Time Internet Services. In *IEEE Computer*, pages 14–16, November 1997.

[PC76] C. Pang and H. Chen. Optimal short-term unit commitment. *Transactions on Power Apparatus and System*, 1976.

[RSA97] 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.

[SPL98] S-Plus5, "Guide to Statistics", MathSoft Inc. September 1988.

[Sr95] A. Srinivasan. Improved approximations of packing and covering problems. In *Proc. ACM Symposium on the Theory of Computing*, pages 268–276, 1995. Complete version appears in *SIAM J. Computing*, 2000.

[Sr97] A. Srinivasan. Improved approximations for edge disjoint paths, unsplittable flow, and related routing problems. In *Proc. IEEE Foundations on Computer Science*, pages 416–425, 1997.

[Sr99] 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.

[We98] Weiss J. Market Power Issues in a Restructuring of the Electricity Industry: An Experimental Investigation. `http://www.ksg.harvard.edu/hepg/Jurgen2.htm.`

[Wi97b] A.M. Wildberger. Issues associated with real time pricing. Unpublished Technical report, Electric Power Research Institute (EPRI), 1997.

[Wi97a] 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.

[Wi97a] A.M. Wildberger. Autonomous adaptive agents for the distributed control of the power grid in a competitive electric power industry. Keynote Address, *IEEE Symposium on Knowledge Engineering and Systems, KES'97*, Adelaide, Australia, 1997.

[Web] 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)

[Wo00] Wolak F. (2000), An Empirical Analysis of the Impact of Hedge Contracts on Bidding Behavior in a Competitive Electricity Market, `http://www.stanford.edu/∼wolak`.

[WSJ] The Wall Street Journal, August 04 2000, pp. A2, August 14 2000 pp. A6 and August 21 2000, November 27 2000, December 13, 14 2000.

[RSP87] J. Rayburn, W. Synder, H. Powell. Dynamic programming approach to unit commitment. *IEEE Transactions on Power Systems*, 1987.

[WW96] A.J. Wood and B.F. Wollenberg. *Power Generation, Operation and Control.* John Wiley and Sons, 1996.

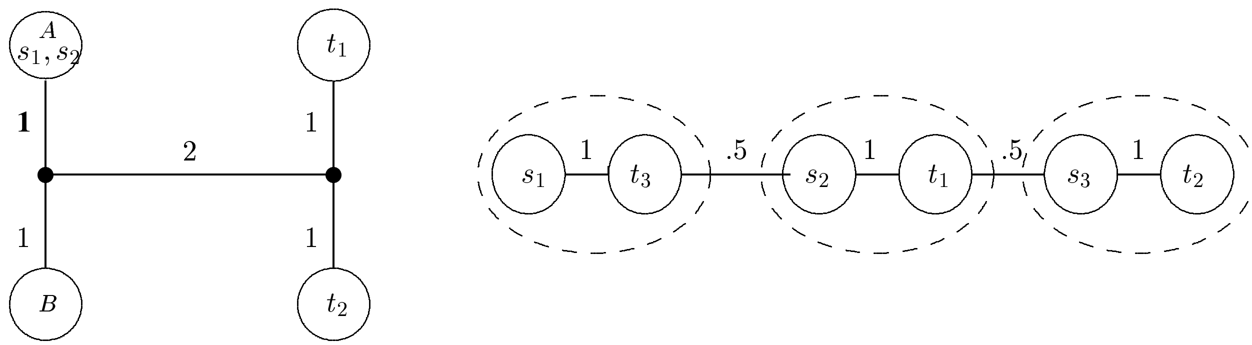
[ZG88] F. Zhuang and F. Galiana. Towards a more rigorous and practical unit commitment by Lagrangian relaxation. *IEEE Transactions on Power Systems*, 1988.

Appendix

A Illustrative Examples

Some unusual characteristics of our problems are now described, following [CF+97].

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, and two consumers. Let us assume that each consumer has a demand of 1 unit. Before deregulation, say both A and B are owned by the same company. If we assume that the plants 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



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

Figure 1: Figures for Examples 1 and 2.

contracts with A. It is clear that both the consumers now cannot get power by A alone. 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*.

B Worst-Case Examples

The three heuristic methods of § 3 can be shown to have worst case performance that is $\Omega(n)$. Example 4 shows that all the heuristics can perform poorly w.r.t. an optimal solution. This is not too surprising given that the optimal solution gets to look at all of the input before clearing the contracts. The next set of examples are designed to show that there are instances when one solution strategy beats the other two, implying that there is no dominating solution strategy.

Example 4: Consider a network with two nodes A and B joined by an edge (A, B). The capacity of the edge (A, B) is C, where $1 \leq C \leq \min\{n/(n-2), 2\}$. There are n (assume n is even) contracts $(s_1, t_1), \ldots, (s_n, t_n)$. Odd numbered Contracts have demand of 1 unit and the sources and sinks of these contracts are distributed as follows: Node sources $s_1, s_3, \ldots s_{n-1}$ are located at node A and their corresponding consumers $t_1, t_3, \ldots t_{n-1}$ are located at B. Let us call this set *Odd-Set.* For the even numbered contracts (denoted *e*ven-set) we have a demand of $1 + \frac{2C}{n}$ per contract and the source sink locations are reversed: the sources are located at B and the sinks at A. Note that (i) All *Odd-set* contracts have demand that is less than every contract in *Even-Set.* (ii) In the absence of any other contracts, only one *Odd-set* contract can be cleared (wlog we say it is (s_1, t_1)). (iii) Similarly, exactly one *Even-set* contract can be cleared if there are no *Odd-set* contracts.

Now consider how many contracts can be satisfied by the each of three heuristic methods. (i) The SMALLEST-FIRST HEURISTIC will clear only one contract (s_1, t_1). (ii) The LARGEST-FIRST HEURISTIC will also clear exactly one contract (s_2, t_2). (iii) RANDOM-ORDER HEURISTIC will also perform poorly with high probability. This is because, of the $n!$ ways to arrange the contracts, it can be shown that only a negligible fraction of the orders are good. (iv) An optimal solution can clear all the contracts simultaneously, since the flows from *Odd-set* appropriately cancel the flows from *Even-Set*. Thus performance guarantee of the SMALLEST-FIRST HEURISTIC and LARGEST-FIRST HEURISTIC is $\Omega(n)$. The performance guarantee of RANDOM-ORDER HEURISTIC is also $\Omega(n)$ with high probability.

Example 5: Again, we have a single edge as our network. Denote the edge by (A, B) as before with the endpoints being A and B respectively and the edge capacity being 1 unit. We have n contracts. As before we divide them into *Even-Set* and *Odd-set* of contracts. The contracts' demands are increasing: the i^{th} contract has demand $1 + (i-1)\epsilon$, where ϵ is chosen so that $0 < \epsilon < 1$ and $(n-1)\epsilon > 1$. It is clear that SMALLEST-FIRST HEURISTIC can clear all the contracts, while LARGEST-FIRST HEURISTIC can clear exactly one contract. Again simple calculation shows that RANDOM-ORDER HEURISTIC will perform poorly with high probability.

C Details of the ILP rounding

We now present details of the ILP rounding procedure, following [CF+97].

(MFP) Maximize $\sum_{i \in R} x_i$

subject to

$$
\begin{array}{rclll}
\mathcal{N}\mathbf{x} & \leq & b & & (1) \\
x_i & \in & \{0,1\} & \forall i \in R & (2) \\
z_{i,e} & \geq & 0 & \forall i \in R,\ e \in E & (3) \\
\sum_i z_{i,e} & \leq & c_e & \forall e \in E & (4) \\
z_{i,e} & \leq & c_e x_i & \forall e \in E, i \in R & (5)
\end{array}
$$

Linear program formulation: Let $N^+(v)$ and $N^-(v)$ denote the set of directed outgoing and incoming edges incident on v respectively. We have two indicator variables x_i for each contract i between pairs (s_i, t_i), and indicator variables $z_{i,e}$ for each (s_i, t_i) pair and each edge e. The intended meaning of x_i is that the total flow for (s_i, t_i) is $d_i x_i$; and the meaning of $z_{i,e}$ is that the flow due to the contract between (s_i, t_i) on edge e is $z_{i,e}$. We will sometimes use $z_{i,\cdot}$ to denote the flow associated with the pair (s_i, t_i). Also recall that we assume that the source sink pairs for each contract are distinct. In the integral version of the problem, we will have $x_i \in \{0, 1\}$.

Constraint (1) is just shorthand for the usual flow constraints for all (s_i, t_i) and says that (i)The total incoming flow $z_{i,e}$ into t_i is $d_i x_i$, and the total outgoing flow $z_{i,e}$ from s_i is also $d_i x_i$. This in particular relates the variables $z_{i,e}$ to x_i. It states that $\sum_{(e) \in N^+(s_i)} z_{i,e} - \sum_{(e) \in N^-(s_i)} z_{i,e} = d_i x_i \quad \forall i \in R$ and $\sum_{(e) \in N^-(t_i)} z_{i,e} - \sum_{(e) \in N^+(t_i)} z_{i,e} = d_i x_i \quad \forall i \in R$ (ii) for all vertices v not in $\{s_i, t_i\}$, the incoming flow of type $z_{i,\cdot}$ i.e. flow due to the contract pair (s_i, t_i) equals the outgoing flow.

Constraint (2) imposes the 0/1-contract satisfaction rule. Constraint (3) says that the flows are non-negative. Constraint (4) says that the total flow corresponding to all source-sink pairs on an edge does not exceed its capacity. Finally, constraint (5) says that the flow due to any particular contract over an edge e does not exceed the fractional capacity of that edge w.r.t. the 0/1-contract satisfaction of that contract.

Performance Guarantee: We briefly discuss the performance guarantee of the above algorithm for the single source case. Let E_e be the "bad event" that edge e's capacity is violated. i.e., that $E_e := Pr(\sum_{i:Y_i=1} (1 -$

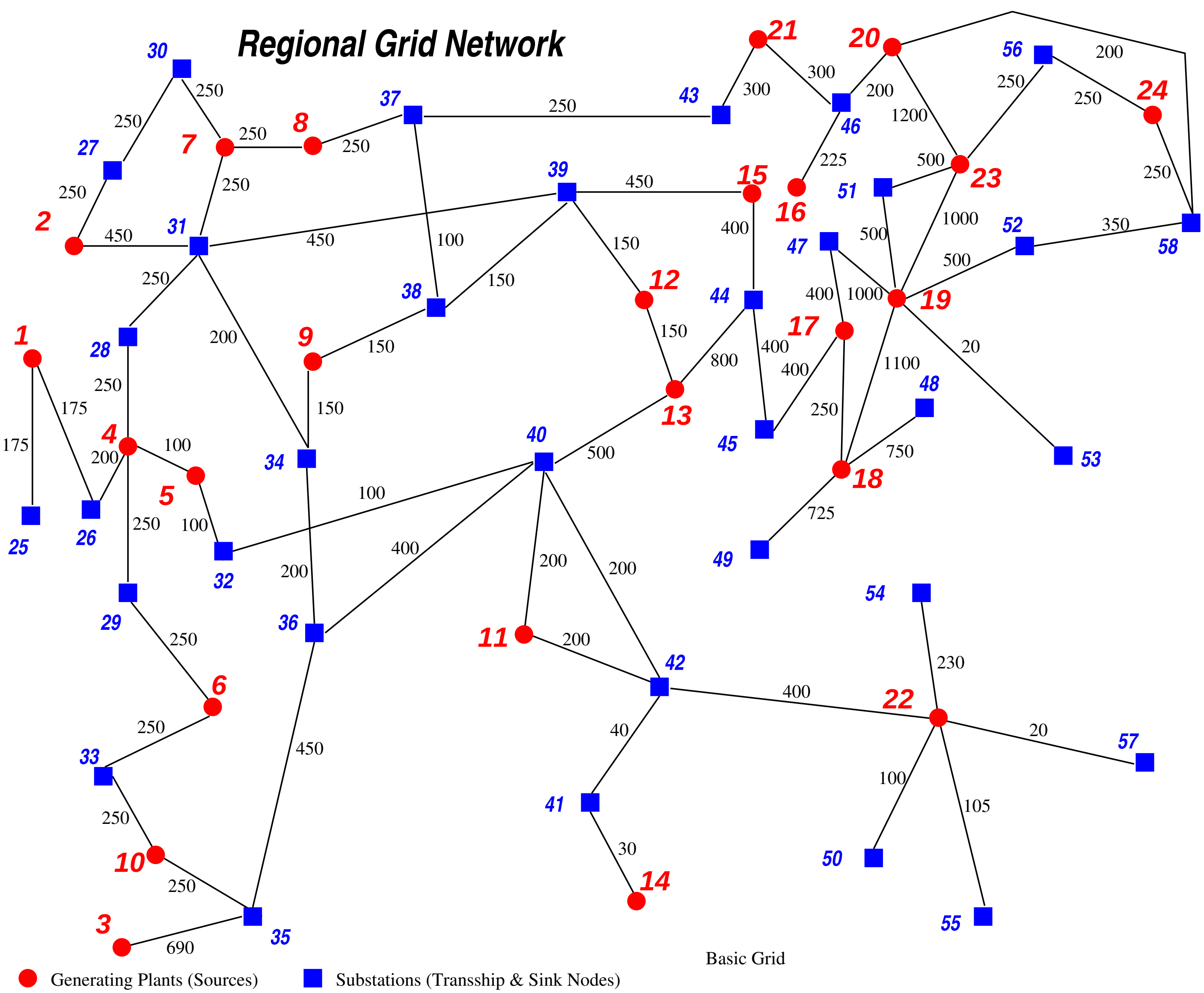


Figure 2: This shows the network with node numbered as they are referenced in all scenarios and edge capacities labeled at values used for **Scenarios 1 & 2**. The placement of the nodes and edges are what is probably the final form. The least number of edges cross and the nodes in the upper right are spread out a little bit maintaining the general feel of the distribution while allowing easier reading.

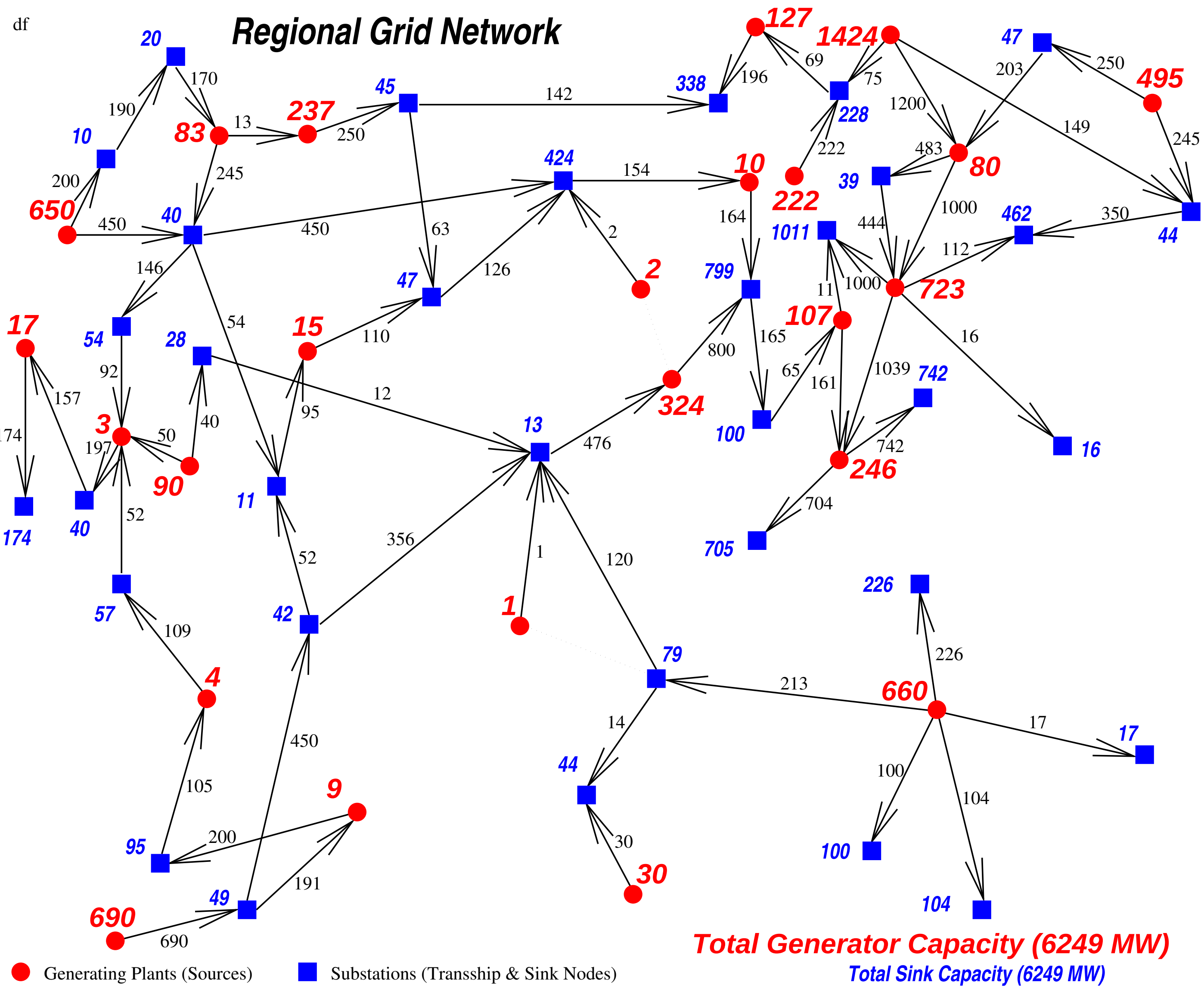


Figure 3: Shows the network with the maximum capacities of the nodes and edges. The arrows indicate the direction of the flow and the numbers associated with the edges show the flow values (not the capacities). The edges with no flow have been changed to dotted lines although one or two of the dotted lines may look solid.

$\epsilon)z_{i,e}/x_e) > c_e$. Let F be the "bad event" that the objective function becomes really small: i.e., $F := Pr(\sum_i w_i Y_i) \leq y^*/(10\lambda)$ where y^* is the LP optimal value, $\sum_i w_i x_i$. Define $\mathcal{E} := (\bigvee_{e \in E} E_e) \vee F$. It is clear that $Pr(\mathcal{E}) \vee Pr(\overline{\mathcal{E}}) = 1$. But $\overline{\mathcal{E}} = \bigwedge_{e \in E} \overline{E_e} \bigwedge \overline{F}$, i.e. $\overline{\mathcal{E}}$ represents the event that (i) No edge capacity is violated too much **and** (ii) the objective function does *not* become small. Thus, we want to show that for a suitable $\lambda > 1$,

$$Pr(\mathcal{E}) = Pr(\bigvee_{e \in E} E_e \bigvee F) < 1, \tag{2}$$

which would imply that $Pr(\overline{\mathcal{E}}) = 1 - Pr(\mathcal{E}) > 0$. We would have thus demonstrated that there exists one point in the space of all integral rounded solutions that is *good*. We will show that this is indeed the case by using the FKG inequality. In particular, we will show that the events $\overline{E_e}$ are *positively correlated* and satisfy the conditions in the FKG inequality. Specifically, we show that

$$\forall e \in E \quad Pr(\bigwedge_{e \in E} \overline{E_e}) \geq \prod_{e \in E} Pr(\overline{E_e}). \tag{3}$$

Using this it suffices to show that

$$Pr(\bigvee_e E_e \bigvee F) \leq Pr(\bigvee_e E_e) + Pr(F) \leq 1 - (\prod_e (1 - Pr(E_e))) + Pr(F) < 1. \tag{4}$$

Each $Pr(E_e)$ and $Pr(F)$ can be upper-bounded by a Chernoff bound. By following the proof of Theorem 2.2 in [Sr97] and also using the ideas in [Sr95] packing/covering, if we take λ to be $c(m/y^*)^{(1-\epsilon)/\epsilon}$ for a suitably large constant c such as 10, then we can show that $(*)$ is true; we can also compute the rounding in deterministic polynomial time. Here are the ideas in brief. Basically, we can upper-bound $Pr(E_e)$ as follows. We want to upper-bound

$$Pr(\sum_{i:Y_i=1} (1-\epsilon)z_{i,e}/x_e) > c_e = Pr(\sum_i (z_{i,e}/(c_e x_i))Y_i) > 1/(1-\epsilon).$$

(which is true since the summation is done over i.) From constraints (4) and linearity of expectations, we get $E[\sum_i (z_{i,e}/(c_e x_i))Y_i] = \frac{x_i}{\lambda}\frac{z_{i,e}}{c_e x_e} \leq 1/\lambda$. From constraints (5), $z_{i,e}/(c_e x_i) \leq 1$. Thus, we have a sum S of independent random variables, each lying in $[0,1]$; $E[S] \leq 1/\lambda$. We want to upper-bound $Pr(S > 1/(1-\epsilon))$; by standard Chernoff analysis, this is at most $O(1/\lambda^{1/(1-\epsilon)})$. Thus, for all e, $Pr(E_e) \leq O(1/\lambda^{1/(1-\epsilon)})$ Similarly, $Pr(F)$ can be upper-bounded by $e^{-\Theta(y*/\lambda)}$, by Chernoff. Plugging these in Equation 4, we can verify that if $\lambda = c(m/y*)^{(1-\epsilon)/\epsilon}$ for a sufficiently large constant c, then Equation 4 holds.

D Statistical Tests

D.1 Contrasts

The next question of interest is what really caused the rejection of the null hypothesis. What are the sources of significant differences that gave rise to the overall significance. Just knowing that at least one of the algorithms is different does not help us identify which algorithm is significantly different. To answer this we use a procedure called *contrast*. A contrast $\mathcal{C}$ among the I population means (μ_i) is a linear combination of the kind: $\mathcal{C} = \Sigma_i \alpha_i \mu_i = \alpha_1\mu_1 + \alpha_2\mu_2 + \cdots + \alpha_I\mu_I$ such that the sum of contrast coefficients, $\Sigma_i \alpha_i = 0$. In the absence of true population means, we use the unbiased sample means which gives the estimated contrast as: $\hat{\mathcal{C}} = \Sigma_i \alpha_i \overline{X}_i = \alpha_1\overline{X}_1 + \alpha_2\overline{X}_2 + \cdots + \alpha_I\overline{X}_I$ The contrast coefficients, α_1, α_2, $\cdots$, α_I are just positive and negative numbers that define the particular hypothesis to be tested. The null hypothesis states that the value of parameter for every contrast is zero, i.e., $H_0 : \mathcal{C} = 0$. The value of the contrast is tested by a F-test to see if the observed value of the contrast is significantly different from the hypothesized value of zero.

From Table 1, it is clear that the randomized rounding algorithm (RR) is different from all the other algorithms for all four scenarios. On an average, RR algorithm satisfies 49% less contracts than the Largest-First (LF) heuristic and 50% less than the Random-Order (RO) heuristic and 52% less contracts than the

Smallest-First (SF) heuristic. The difference between SF, RO and LF heuristics appears only marginal. Based on this observation, we constructed the following contrast that tests if RR is statistically significantly different from the other three algorithms: $\mathcal{C}_{1Q} = \frac{1}{3}(\overline{X}_{2\cdot\cdot}) + \frac{1}{3}(\overline{X}_{3\cdot\cdot}) + \frac{1}{3}(\overline{X}_{4\cdot\cdot}) - \overline{X}_{1\cdot\cdot}$

Using the value of algorithm means from Table 1 we can calculate the value of $\mathcal{C}_{1Q}$ (Q stands for the quality of solution) to be equal to 0.50.[4] The sum of squares of a contrast is expressed as: $SS(\mathcal{C}_{1Q}) = \frac{(\mathcal{C}_{1Q})^2}{\Sigma_i \alpha_i^2 / N_i}$ Here α_i are the coefficients of the contrast and $N_i = 120$ is the number of observations (i.e. sample points for each algorithm across all scenarios). This results in $SS(\mathcal{C}_{1Q}) = 22.68$. Now we can use the following F-test to see the significance of the contrast: $SS(\mathcal{C}_{1Q})/MSE \sim F(1, 464)$

MSE stands for the mean square error of the residuals. The contrast has one degree of freedom and residuals have 464 degrees of freedom (see table 3). F = 22.68/.0008 = 28350, since the observed value of F-test is greater than the critical F-value given in the F-distribution table, for any significance level, the null hypothesis is rejected. This confirms our earlier observation that the RR algorithm is significantly inferior in performance compared to the other three algorithms. The sum of squares of $SS(\mathcal{C}_{1Q}) = 22.68$ shows that 98% of the variation in factors sum of squares (total factors sum of squares being 23.05 i.e. total SS - residual SS, see table 1) is due to the difference in RR algorithm versus the other three algorithms.

Table 1 shows that the first three scenarios clear about 86% of the optimal number of contracts while under the fourth scenario, the number of contracts cleared is less than 82% of the optimal. Even though the difference in the number of cleared contracts is not very big, one would be curious to find out if the difference in performance under the first three scenarios versus the fourth scenario is significant or not. To answer this we created the following contrast which is orthogonal[5] to the first contrast ($\mathcal{C}_{1Q}$):

$\mathcal{C}_{2Q} = \frac{1}{3}(\overline{X}_{\cdot 1 \cdot}) + \frac{1}{3}(\overline{X}_{\cdot 2 \cdot}) + \frac{1}{3}(\overline{X}_{\cdot 3 \cdot}) - \overline{X}_{\cdot 4 \cdot}$. Just like $\mathcal{C}_{1Q}$, we can calculate the value of $\mathcal{C}_{2Q}$ using table 1. $SS(\mathcal{C}_{2Q})/MSE \sim F(1, 464) = 0.14/.0008 = 175$ Again, the null hypothesis is rejected implying that the fourth scenario is indeed significantly different from the other three scenarios. Now we look at two more contrast to check if SF and LF are significantly different ($\mathcal{C}_{3Q}$) and LF and RO are significantly different ($\mathcal{C}_{4Q}$).

$$\mathcal{C}_{3Q} = \overline{X}_{2\cdot\cdot} - \overline{X}_{3\cdot\cdot}, \mathcal{C}_{4Q} = \overline{X}_{3\cdot\cdot} - \overline{X}_{4\cdot\cdot}$$

$$SS(\mathcal{C}_{3Q})/MSE \sim F(1, 464) = 2.178/.0008 = 2722.5$$

$$SS(\mathcal{C}_{4Q})/MSE \sim F(1, 464) = 1.038/.0008 = 1297.5$$

For both $\mathcal{C}_{3Q}$ and $\mathcal{C}_{4Q}$, the observed value of the F-test is greater than the critical F-value given in the F-distribution table, the null hypothesis in both cases are rejected implying that, SF provides a better solution than LF and also, RO performs significantly better than LF.

In summary, all algorithms show significantly different performance when measured in terms of quality of solution. On an average, the best solution is given by the SF heuristic and the worst by the RR.

D.2 Running-Time

Tables 4 and 2 show results of the same experiment when performance is measured by the *running time* of the algorithm. The factors, number of observations, kinds of tests, everything remained the same as before except the performance measure. Table 3 results clearly demonstrate that different algorithms take significantly different time to run and different scenarios have significantly different running time. The interaction term is significant at any level of confidence implying that the running time of an algorithm is different for different scenarios.

Table 2 shows that the RR algorithm takes noticeably more time to run as compared to the other three heuristics. Among the three heuristics, LF and RO take about the same time but SF takes about 19 megaticks

[4] The table values are shown in percentages, but here we use actual values.

[5] Two contrasts $\mathcal{C}_1$ and $\mathcal{C}_2$ are said to be orthogonal if the sum of the products of their corresponding coefficients is zero. It is desirable to have independent or orthogonal contrasts because independent tests of hypotheses can be made by comparing the mean square of each such contrast with the mean square of the residuals in the experiment. Each contrast has only one degree of freedom.

Source	DF	SS	MS	F**-test**	p**-value**
Scenario (Block)	3	21152	7050.8	56.97	0
Algorithm (Treatment)	3	2161199	720399.8	5821.07	0
Scenario:treatment	9	28156	3128.5	47.78	0
Residuals	464	30381	65.5		
Total	479	2240888			

Table 4: **Results of Two-Factor ANOVA:** This table shows results of two-factor ANOVA where the factors are algorithms and scenarios. The measurement is the running time of the algorithm. The p-values show that the algorithm effect, scenario effect and the interaction between the algorithms and scenarios are all significant at any level of confidence.

more than the LF and RO. Similarly, scenario 3 and 4 take about the same time but scenario 1 and 2 look different. To test all the above mentioned observations, we create the following different contrasts: $\mathcal{C}_{1t} = \frac{1}{3}(\overline{X}_{2..}) + \frac{1}{3}(\overline{X}_{3..}) + \frac{1}{3}(\overline{X}_{4..}) - \overline{X}_{1..}$, $\mathcal{C}_{2t} = \frac{1}{2}(\overline{X}_{3..}) + \frac{1}{2}(\overline{X}_{4..}) - \overline{X}_{2..}$, $\mathcal{C}_{3t} = \overline{X}_{3..} - \overline{X}_{4..}$, $\mathcal{C}_{4t} = \overline{X}_{.1.} - \overline{X}_{.2.}$

All the above contrasts are orthogonal to each other. The first contrast, $\mathcal{C}_{1t}$ (here t stands for running time), checks if the RR algorithm takes more time to run than the other three heuristics. The second contrast, $\mathcal{C}_{2t}$, will find if the SF heuristic is significantly different from the LF and RO heuristic. The third contrast, $\mathcal{C}_{3t}$, checks if the LF and RO heuristics take about the same time to run. Finally, contrast $\mathcal{C}_{4t}$, check if the first scenario takes less time to run than the second scenario. The results of all the contrasts are shown below.

$$SS(\mathcal{C}_{1t})/MSE \sim F(1, 464) = 2133700.9/65.5 = 32575.5$$

$$SS(\mathcal{C}_{2t})/MSE \sim F(1, 464) = 27424.4/65.5 = 418.69$$

$$SS(\mathcal{C}_{3t})/MSE \sim F(1, 464) = 72.6/65.5 = 1.11$$

$$SS(\mathcal{C}_{4t})/MSE \sim F(1, 464) = 20093.4/65.5 = 306.76$$

As can be seen by looking at the F-distribution table, all the above contrasts except $\mathcal{C}_{3t}$ show that the observed value of the F-test is greater than the critical F-value. Hence the null hypothesis i.e. $H_0 : \mathcal{C}_{it} = 0$ where $i = 1, 2, 4$ can be rejected at any level of significance. This confirms our earlier hypothesis that RR indeed takes longer to run than the other three heuristics. SF takes more time to run than the LF and RO heuristics and the second scenario takes significantly more time to run than the first scenario.

The mean difference in running time across different algorithms shows that all algorithms are significantly different in terms of running time except for the *Largest-First* and the *Random-Order* heuristics. These two heuristics take about the same time to run and indeed a contrast done i.e. $\mathcal{C}_{3t}$ on LF and RO proves that and the null hypothesis, $H_0 : \mathcal{C}_{3t} = 0$, is accepted.

The randomized rounding algorithm takes significantly more time to run than any of the other heuristics. The gap in running time narrows when RR is compared against SF. RR takes 141 megaticks more time than the SF heuristic, 160 megaticks more than the LF and RO. SF takes more time to run than LF and RO but it clears more contracts than LF and RO.