94 WM 118-0 EC

Integrating Price-Based Resources In Short-Term Scheduling Of Electric Power Systems

Alva J. Svoboda and Shmuel S. Oren
Department of Industrial Engineering and Operations Research
University of California, Berkeley
Berkeley, CA 94720

Abstract--**This paper discusses the application of a Lagrangian relaxation algorithm for solving short-term resource scheduling problems in the presence of "endogenously priced" resources. We demonstrate how such resources can be incorporated into the scheduling model as opposed to the prevailing practice of activating them through post-dispatch price signals derived from the short term scheduling of resources directly controlled by the utility . Through a simulation study we compare operations and costs of integrated scheduling to those resulting from the post-dispatch approach. Our analysis demonstrates that integrated scheduling can produce significant improvements.**

I. INTRODUCTION

Electric power systems both in the US and internationally are increasingly characterized by diverse and complex resource types and rapid evolution of system operating constraints. In particular, the proportion of many systems' resource mix filled by so-called "endogenously priced" [6] resources is growing, in some cases quickly. These resources, which include dispatchable independent power production, load management and interruptible customers, economy transactions with neighboring systems, and storage resources such as pumped-storage hydro plants, share in common the attribute of not having a well-defined a priori operating cost. Rather, their operating costs are functions of the operations of the system as a whole. For example, such resources may be dispatched based on the assumption that the system marginal operating costs determine the prices at which the system can buy and sell energy on the margin.

Lagrangian relaxation-based methods of resource scheduling, it will be argued here, are well-suited for application in power systems with significant amounts of endogenously priced generation. The use of such methods has become prevalent in recent years, at least for large and complex systems, because they have allowed more detailed and flexible representation of resource and system characteristics than would be possible using other available methods of scheduling. But applications of Lagrangian relaxation to resource scheduling have typically not integrated endogenously priced resources into the scheduling model. Rather, marginal cost information obtained from an entirely "supply-side" (i.e., directly controlled by the utility) resource schedule is used as a first-order approximation of the marginal value of additional generation or load relief. This approach is often referred to as "post-dispatch" scheduling of price-based resources.

94 WM 118-0 EC A paper recommended and approved by the IEEE Energy Development and Power Generation Committee of the IEEE Power Engineering Society for presentation at the IEEE/PES 1994 Winter Meeting, New York, New York, January 30 - February 3, 1994. Manuscript submitted July 9, 1993; made available for printing January 11, 1994.

This paper will attempt to demonstrate that endogenously priced resources can in fact be incorporated into Lagrangian relaxation resource scheduling models. It will show, further, that the "second-order" effects of the presence of endogenously priced resources in a system can be significant in systems of realistic size and diversity. Finally, it will argue that Lagrangian relaxation provides a useful framework for negotiating the "value of dispatchability" between a utility (or central system operator) and endogenously priced resources contracting with the utility to provide generation, subject to certain operating constraints.

The following sections of this paper will briefly present a formulation of the resource scheduling problem and a description of a Lagrangian relaxation algorithm for solving it. We will then present an example of an application of Lagrangian relaxation to resource scheduling in the presence of endogenously priced resources. We will compare operations and costs for the price signals obtained from integrated scheduling of all resources to price signals obtained from scheduling of only the resources directly controlled by the utility. The latter set of price signals correspond to the marginal energy costs obtained from a conventional economic dispatch.

II. FORMULATION AND SOLUTION OF THE RESOURCE SCHEDULING PROBLEM

The suitability of Lagrangian relaxation for the scheduling of endogenously priced resources is a by-product of the structure of the Lagrangian relaxation algorithm, which solves a Lagrangian dual of the resource scheduling problem: the dual variables act as "price signals" in the algorithm to yield a set of resource schedules in which each resource maximizes the profits of its own operations based on the prices it is offered. When the set of resource schedules is close to satisfying system requirements (such as the meeting of customer demand, or the requirement for spinning reserves), the algorithm ensures that these schedules will also be near-optimal in terms of minimizing total production costs. The dual multipliers corresponding to the near-feasible, near-optimal set of resource schedules are thus near-optimal price signals (at least in theory), and could be communicated to resources not under direct utility control in order to obtain the desired resource operations.

The resource scheduling problem is to find an efficient and feasible resource commitment and forecasted economic dispatch for a given power system and forecasted demands over a given scheduling horizon. The formulation given here assumes that all system operating costs associated with commitment and dispatch can be attributed to individual resources in the system. It further assumes that resource operating costs take the form of either generation costs, or "change-of-state" costs such as startup costs.

The decision variables in the mathematical program representing the resource scheduling problem consist of commitment decisions represented as changes of resource state, and dispatch decisions representing average resource generation levels. Constraints on the resource commitment and dispatch decisions are distinguished according to whether they apply to individual resources only, or to multiple resources at once. The latter are termed "coupling constraints." Constraints applying only to individual resources may take arbitrary form.

The formulation of the resource scheduling problem to be investigated in this paper has as its only coupling constraint the satisfaction of energy demand in each subperiod of the scheduling horizon. The problem may be written as follows:

$$\underset{\{x_{it},p_{it}\}}{Minimize} \sum_{t=1}^{T}\sum_{i=1}^{I}[c_{it}(x_{(i,t-1)},x_{it}) + f_{it}(p_{it})]$$

$$subject\ to\ \sum_{i=1}^{I} p_{it} \geq D_t,\ t=1,\dots,T \qquad (1)$$

$$\bar{x}_i \in \bar{X}_i,\ i=1,\dots,I$$

$$p_{it}^{\min}u_i(x_{it}) \leq p_{it} \leq p_{it}^{\max}u_i(x_{it}), i=1,\dots,I, t=1,\dots,T \quad (2)$$

In this formulation, the i'th resource's commitment state in period t is denoted by x_{it}, and its generation level is denoted by p_{it}. The costs to be minimized in the objective function are the change of state costs, denoted by $c_{it}(x_{(i,t-1)},x_{it})$, and the costs of generation, denoted by $f_{it}(p_{it})$ for generation level p_{it}.

Constraint set (1) represents the supply-demand balance requirements applying in each of the T periods of the scheduling horizon. Typically these are written as equality constraints, but they have been written as inequality constraints because the cost minimization objective function implies that an optimal solution will satisfy these constraints with equality where possible.

Constraint set (2) represents additional constraints on individual resource schedules over the scheduling horizon. $u_i(x_{it})$ is a function which gives the fraction of the i'th resource's capacity considered to be available given the resource's commitment state. Thus, $0 \leq u_i(x_{it}) \leq 1$, and in cases where the commitment state is either "off" or "on," $u_i(x_{it}) \in \{0,1\}$. Constraints on $\bar{x}_i = (x_{i1}, x_{i2}, \dots, x_{iT})$, the trajectory of commitment states, may be arbitrarily complicated, depending on the i'th resource's operating characteristics. $\bar{X}_i$ may for example represent all allowed paths in a dynamic program's state-transition network. The resource generation levels are assumed to be constrained between time-dependent minimums and maximums when resources are committed, and constrained to be zero when resources are not committed.

The use of Lagrangian relaxation to solve the resource scheduling problem was described in [9] as early as 1977. Initially, the Lagrangian relaxation was proposed for use within a branch-and-bound algorithm to solve a mixed-integer formulation of the problem exactly, but it was soon observed that the Lagrangian relaxation yielded a near-optimal solution at the very beginning of the branch-and-bound procedure, and that branch-and-bound "fathoming" did not yield improvements in the solution significant enough to justify the increased computational time required. It was further demonstrated [1] that the quality of the solution yielded by Lagrangian relaxation actually improves with increases in the size of the scheduling problem, where size is measured in terms of the number of non-identical resources and the number of periods in the scheduling horizon.

The degree of detail in the system representation allowed by Lagrangian relaxation implies potentially very large input and output data sets in practical applications, but advances in computer memory and database technology have made such applications more feasible over the years since the technique was first proposed for the resource scheduling problem. Electricite de France developed several applications of the method [8]. Applications intended for general use by power system planners have been developed by Decision Focus, Inc. [4]. Pacific Gas and Electric has developed a more specialized application which includes detailed modeling of its interconnected hydro resources and its large pumped-storage plant [5].

In Lagrangian relaxation algorithms, solution of the resource scheduling problem is based on maximizing a Lagrangian dual of the problem. For the formulation given above, the dual takes the form

$$\underset{\bar{\lambda}}{Maximize}\ q(\bar{\lambda}) \quad subject\ to\ \lambda_t \geq 0,\ t=1,\dots,T$$

$$\text{where } q(\bar{\lambda}) = \underset{\{x_{it},p_{it}\}}{Min} \sum_{t=1}^{T}\sum_{i=1}^{I}[c_{it}(x_{(i,t-1)},x_{it}) + f_{it}(p_{it})] + \sum_{t=1}^{T}[\lambda_t(D_t - \sum_{i=1}^{I} p_{it})]$$

$$subject\ to\ \bar{x}_i \in \bar{X}_i,\ i=1,..$$

$$p_{it}^{\min}u_i(x_{it}) \leq p_{it} \leq p_{it}^{\max}u_i(x_{it}),\ i=1,\dots,I,\ t=1,\dots,T.$$

The dual function $q(\overline{\lambda})$ is separable in the contributions to the dual objective of the individual resources, and thus can be written

$$q(\overline{\lambda}) = \sum_{i=1}^{I} q_i(\overline{\lambda}) + \sum_{t=1}^{T} \lambda_t D_t$$

where

$$q_i(\overline{\lambda}) \quad \underset{(x_{it},p_{it})}{Min} \sum_{t=1}^{T} [c_{it}(x_{(i,t-1)}, x_{it} \; + f_{it}(p_{it})] \quad \sum_{t=1}^{T} [\lambda_t p_{it}]$$

$$subject \;\; to \;\; \overline{x}_i \in \overline{X}_i,$$

$$p_{it}^{\min} u_i(x_{it}) \leq p_{it} \leq p_{it}^{\max} u_i(x_{it}), \;\; t \quad 1,\ldots,T.$$

Each $q_i(\overline{\lambda})$ is evaluated by solving a subproblem involving only the i'th resource. This subproblem may be interpreted as the i'th resource's profit maximization when it sees the price vectors $\overline{\lambda}$ for its hourly generation. The vector of optimal dual multipliers $\overline{\lambda}^*$ has been interpreted as the marginal values to the system of resource generation, or the marginal energy production costs, in each hour. Since there are I resources, I resource subproblems must be solved to evaluate $q(\overline{\lambda})$ for particular $\overline{\lambda}$.

Lagrangian relaxation algorithms maximize the dual iteratively. On each iteration, the multipliers ($\overline{\lambda}$) are updated and $q(\overline{\lambda})$ reevaluated by solving the resource subproblems given the new multiplier values.

One simple formula for updating the multipliers makes use of the fact that a subgradient of the dual objective function $q(\overline{\lambda})$ can be formed as a vector of the differences between the right-hand and left-hand sides of the coupling constraints. Thus, for example the vector $g = (g_1, \ldots g_r)$ where, $g_j = D_j - \sum_{i=1}^{I} p_{ij}$, is a subgradient of the dual objective function $q(\overline{\lambda})$. $q(\overline{\lambda})$ is concave in $\overline{\lambda}$, and therefore as demonstrated in [10] an update on the k'th iteration of the form

$$\overline{\lambda}^k = \overline{\lambda}^{k-1} + \beta^k g$$

converges when the $\{ \beta^k \}$ are a (possibly predetermined) sequence of scalars such that $\beta^k \to 0$ as $k \to \infty$ and $\sum^{\infty} \beta^k = \infty$

III. Uses of the Dual Solution

Unfortunately, because the scheduling problem as formulated above is a mixed-integer nonlinear program with an objective function that is likely to be non-concave, an optimal solution to the dual problem does not immediately yield an optimal and feasible solution to the scheduling problem itself. That would only be the case if the optimal value of the primal objective function equaled the optimal value of the dual objective function, so that the duality gap between the two solutions was zero.

Because the dual objective function value must always be less than or equal to the optimal value of the primal objective function, the fact that the resource schedules $\{(x_{it}^*, p_{it}^*)\}$corresponding to the optimal dual solution $\{\overline{\lambda}^*\}$are guaranteed by the procedure to be feasible to the individual resources' operating constraints means that $\{(x_{it}^*, p_{it}^*)\}$will be a "better-than-optimal" solution, in terms of cost, which does not satisfy all of the system requirements.

The relatively small size of the duality gap between the near-optimal solution of the dual problem and a nearby feasible solution of the primal scheduling problem itself is a result of several apparent structural characteristics of the problem. The daily and weekly cycles displayed by hourly system loads ensure that a fairly limited subset of all possible resource scheduling decisions are candidates for a good schedule. The high positive correlation between commitment costs and resource size, and the high negative correlation between commitment costs and variable operating costs of resources, ensure that resources not subject to energy limits divide naturally into "base loaded" (or weekly cycling), "cycling" (i.e., shutting down and starting up again at least once over the scheduling horizon), and "peaking" categories, and that there is only limited substitutability of resources between categories. On the other hand, substitutability of resources within categories leads to the possibility of large numbers of near-optimal schedules with similar total costs but different resource commitments, so that the dual objective function is likely to be quite flat near its optimum.

Thus, the dual maximization in Lagrangian relaxation almost always yields a resource commitment schedule that is infeasible with respect to one or more constraints in the original scheduling problem. When the system energy demand requirements are formulated as inequalities, as above, infeasibility must mean that the schedule yields insufficient generation in one or more periods. The next step in developing a practical resource scheduling procedure based on Lagrangian relaxation is determining how to get from this infeasible commitment to a feasible one which doesn't increase costs much. This next step, often termed the scheduling algorithm's "feasibility phase," must necessarily proceed along heuristic lines because the dual optimum yields little information about how close or in what direction the nearest feasible schedule is, or even whether the nearest feasible solution is optimal or near-optimal.

However, it can be argued that the solution of the scheduling problem's dual maximization, obtained via the Lagrangian relaxation procedure described above, serves as a useful and reasonable approximate schedule of short-term electric power system operations based on price signals. Recall that the dual solution consists of two parts: a set of Lagrange multipliers which can be interpreted as shadow prices on the system requirements in each period; and a set of resource schedules, satisfying each individual resource's operating constraints and maximizing each resource's net profits assuming that the multipliers are the actual prices offered for generation in each period. Thus, if the fact that the resource schedules do not in general satisfy all system requirements is disregarded, the dual solution offers a schedule for price-based system operations. The utility could, under the scheduling model's assumptions, communicate a set of prices for generation over the scheduling horizon, and get in response the resource schedules predicted by the dual solution.

The dual solution actually corresponds to a feasible, and therefore optimal (because the duality gap must be zero in this case), solution to a problem having the same structure as the original problem, but with slightly different system requirements. One approach (see e.g. [11]) is to recognize that from any set of multipliers $\{\hat{\lambda}_t\}$ a set of resource schedules $\{\hat{x}_{it}, \hat{p}_{it}\}$ may be derived by solving the individual resource subproblems given these multipliers. Modified system requirements in period t may then be constructed by setting $\hat{D}_t = \sum_{i=1}^{I} \hat{p}_{it}$, $\{\hat{x}_{it}, \hat{p}_{it}\}$ are then optimal schedules solving the modified scheduling problem

$$\underset{\{x_{it}, p_{it}\}}{Minimize} \sum_{t=1}^{T} \sum_{i=1}^{I} [c_{it}(x_{(i,t-1)}, x_{it}) + f_{it}(p_{it})]$$

$$subject\ to \sum_{i=1}^{I} p_{it} \geq \hat{D}_t, \ \ t = 1,...,T$$

$$\overline{x}_i \in \overline{X}_i, \ \ i = 1,...,I,$$

$$p_{it}^{\min} u_i(x_{it}) \leq p_{it} \leq p_{it}^{\max} u_i(x_{it}), \ \ i = 1,...,I, \ \ t = 1,...T$$

The modified scheduling problem corresponding to the optimal dual solution would have right-hand sides $D_t^* = \sum_{i=1}^{I} p_{it}^*$, where $\{p_{it}^*\}$ were the generation schedules derived from the optimal dual multipliers $\{\lambda_t^*\}$. If the difference between the forecasted system requirements $\{D_t\}$ and the modified requirements $\{D_t^*\}$ were small by some measure, then the optimal solution to the modified problem, which would also be the optimal solution to the dual of the original problem, could be taken as a good approximate schedule of operations under the forecasted conditions. Indeed, the magnitude of load forecast errors might well be such that the approximate schedule would be as good in some expected sense as the true optimal solution to the original scheduling problem. In either case, it must be possible to modify the schedule to match actual system conditions, whether by changing the generation levels of already scheduled resources or by using other resources whose primary purpose is to follow the actual load (as pumped storage is used in the English power system's operation).

The argument for the dual solution as a good approximate schedule is a reasonable one if the dual solution has certain characteristics. If the infeasibilities in each period are relatively small then they may be treated as within the range of forecast errors, as mentioned above. If the total "costs of infeasibility" $\sum_{t=1}^{T} \lambda_t (D_t - \sum_{i=1}^{I} p_{it})$ are small then they contribute relatively little to the dual objective function value, which may thus be taken as a good estimate of total electricity production costs at near-optimal efficiency. If these solution characteristics do not hold true in a particular case, then a feasibility phase will be required to get a realistic schedule of system operations. But unfortunately, in this case it may not be possible to find a set of marginal energy prices $\{\lambda_t\}$ that would cause the system resources to schedule themselves according to the feasible solution.

IV. POST-DISPATCH VS. INTEGRATED DISPATCH OPERATIONS

There are two credible approaches to nondiscriminatory price-based operations of power systems. The first, which may be termed the "post-dispatch" approach, starts with a predetermined set of marginal energy prices, based for example on an economic dispatch of utility-controlled resources, and sends these prices out as dispatch signals to all endogenously priced resources. These resources' operations are thus assumed not to affect the values of the system marginal energy costs significantly: they are therefore taken to be the "marginal resources" in all periods. "Post-dispatch" scheduling of endogenously priced resources assumes that the potential benefits of such resources are captured by the marginal production costs in the absence of such resources, and that therefore these marginal costs can serve as the set of inputs to price-based dispatch.

The post-dispatch model of price-based dispatch corresponds to existing utility real-time pricing programs, which set prices to real-time pricing customers a day in advance based on the supply-side resource schedule's marginal production costs. The model also corresponds to the procedure for the dispatch of interruptible customers proposed in [2]. The post-dispatch procedure presumes that endogenously priced resources cannot be treated as dependable in the scheduling process, and moreover that the effects of the dispatch of these resources on meeting total load are adequately reflected in the scheduling model's treatment of load forecast error, e.g., by setting reserve requirements over and above meeting forecasted loads. Underlying these system modeling assumptions is the operational assumption that the system is dispatched economically in real time to satisfy actual demands, but that neither the presence of endogenously priced resources nor the differences between forecasted and actual demands significantly affect marginal production costs.

If an endogenously priced resource receives the marginal energy cost per unit of energy "produced" (interrupted, in the case of demand-side resources) in each period, it receives total payments of $\sum_{t=1}^{T} \lambda_t p_{it}$ over the scheduling horizon. If the resource receives payments (seen by the utility as operating costs) that differ from the marginal energy costs, but the utility dispatches the resource based on "post-dispatch" prices for alternative sources of energy, then the net value of its schedule to the utility may be estimated by the resource subproblem's objective function,

$$q_i(\bar{\lambda}) = -\sum_{t=1}^{T} \lambda_t p_{it} + \sum_{t=1}^{T} [c_{it}(x_{(i,t-1)}, x_{it}) + f_{it}(p_{it})].$$

This value estimates the total reduction in utility costs due to introduction of the resource into the system. This corresponds to the estimate of "dynamic operating benefits" described by Decision Focus, Inc., in an EPRI report (1989). Note that if the resource receives payments based solely on the marginal energy costs themselves, and is scheduled on that basis (as would presumably be the case for real-time pricing customers), its net operating value to the utility by this estimate would be by definition zero, because the marginal payments would equal the marginal benefits of its operations.

The second approach to price-based scheduling, an integrated commitment and dispatch based on Lagrangian relaxation, views the entire set of resources including those endogenously priced as equally schedulable. To estimate the value of a resource to the utility, the scheduling problem may be solved for a base case without the resource, yielding solution $\{\lambda_t^1, x_t^1, p_t^1\}$, and a change case with the resource yielding solution $\{\lambda_t^2, x_t^2, p_t^2\}$, where the schedules x_t^2, p_t^2 include the schedule of the new resource. The value is then estimated by taking the difference between the dual objective function values of the base case and the change case. The estimated value is:

$$q(\lambda^2) - q(\lambda^1) = \sum_{t=1}^{T} \sum_{i=1}^{I+1} [c_{it}(x_{(i,t1)}^2, x_{it}^2) + f_{it}(p_{it}^2)]$$
$$+ \sum_{t=1}^{T} \lambda_t^2 (D_t - \sum_{i=1}^{I+1} p_{it}^2)$$
$$- \sum_{t=1}^{T} \sum_{i=1}^{I} [c_{it}(x_{(i,t1)}^1, x_{it}^1) + f_{it}(p_{it}^1)] - \sum_{t=1}^{T} \lambda_t^1 (D_t - \sum_{i=1}^{I} p_{it}^1)$$

If the infeasibility cost terms $\sum_{t=1}^{T} \lambda_t^2 (D_t - \sum_{i=1}^{I+1} p_{it}^2)$ and $\sum_{t=1}^{T} \lambda_t^1 (D_t - \sum_{i=1}^{I} p_{it}^1)$ are both small, then this estimate is approximately the reduction in production costs due to the presence of the new resource. If $\hat{q}(\lambda^2) - q(\lambda^1)$ is small,

$$\text{where, } \hat{q}(\lambda^2) = \sum_{t=1}^{T} \sum_{i=1}^{I} [c_{it}(x_{(i,t-1)}^2, x_{it}^2) + f_{it}(p_{it}^2)]$$
$$+ \sum_{t=1}^{T} \lambda_t^2 (D_t - \sum_{i=1}^{I} p_{it}^2),$$

then $q(\lambda^2) - q(\lambda^1)$ is approximately equal to the estimated benefits of introducing the resource under "post-dispatch" operations. One condition under which $\hat{q}(\lambda^2) - q(\lambda^1) = 0$ is that $x_t^2 = x_t^1$, $p_t^2 = p_t^1$, $\lambda_t^2 = \lambda_t^1$ for $t = 1, \ldots, T$, $i = 1, \ldots, I$.

Both the integrated dispatch and post-dispatch estimates of net value just presented are based on nondiscriminatory pricing by the utility. In other words, the payments made to price based resources (which include the net cost to the resource owner and the resource profit) are indistinguishable in the utility's perspective from the operating costs of utility-owned resources. An alternative, nondiscriminatory form of price-based operations would have the utility offer a single set of generation prices $\{\lambda_t\}$ to all endogenously priced resources, but require the resources to provide information about their operating costs and characteristics to the utility so that they could be scheduled as efficiently as possible. Under this alternative form of price-based operations, endogenously priced resources receive the net value of the resource (estimated above) as profit. We will assume this alternative form in simulating integrated dispatch and post-dispatch under price-based operations. In describing the results of the simulation we will refer to the utility savings in production cost resulting from the resource use as "utility profit" while the difference between the price paid by the utility for a resource and its actual cost is referred to as the "resource profit."

V. Simulation Set-up

The two forms of price-based operations described above, integrated dispatch and post-dispatch scheduling, are simulated given endogenously priced resources with varying operating constraints. The main objectives of the simulation have been to demonstrate the feasibility of integrated dispatch, compare the effects of the two approaches on distribution of benefits between the utility and the resource owners and examine the impact of the various operational constraints on these benefits.

In simulating the dispatch of endogenously priced resources, we have employed data for a fictitious system equivalent to the CALECO power system data [7] developed for the purpose of testing production costing simulation programs for the California Public Utilities Commission. However, in our simulations we have removed from the original CALECO system all nondispatchable resources (i.e., base-loaded generation such as nuclear and run-of-river hydro) and the load they serve.

In each case the price-based resource has a capacity of 1000 MW (approximately 10% of the total installed dispatchable capacity), a minimum dispatch level of 100 MW, and an assumed variable operating cost of $25/MWH. The capacity was chosen so the resource might have a more than marginal impact on system operations. The operating cost was chosen based on the marginal operating costs of the system before introducing the resource, to make it operate strictly between its minimum and maximum operating limits during as many peak hours as possible and so give its dispatchability maximum value in this simulation.

In performing the following simulations, we employed portions of the code developed for Pacific Gas and Electric Co.'s short-term scheduling program, the "Hydro-Thermal Optimization" or HTO program. Hydro modeling has been drastically simplified, however, by implementing a simple energy-limited resource model to simulate the dispatch of total pondage hydro. In addition, we incorporated several new resource models into the program, to capture operating constraints (such as maximum up time and maximum number of startups in a limited period) which might be peculiar to endogenously priced resources.

A range of operating characteristics are simulated for the incremental, endogenously priced resource: (1) complete dispatchability (i.e., no operating constraints except for the minimum and maximum operating limits); (2) 16 hours minimum down time (i.e., minimum interval between distinct commitments); (3) a $50,000 fixed, or startup cost, per distinct commitment (this cost equals the operating cost of two hours' operation at full capacity); (4) 8 hours maximum up time (i.e., maximum commitment duration) combined with 16 hours minimum down time; (5) a maximum of 3 distinct commitments allowed over the week, each having maximum duration or up time of 8 hours.

VI. Simulation Results: The Effect of Operating Constraints

Table I tabulates the results of the simulations of system operations for each variety of resource. Production costs are estimated by the value of the dual objective function at algorithm termination, and are compared to the estimated production costs for the base case system to obtain estimated resource profits in absolute dollars and relative percentage terms. The costs of imposing constraints on the operations of the endogenously priced resources are estimated by comparing the profits with constraints to the profits under complete dispatchability, and these costs again are given in both absolute and relative terms. Each resource configuration was simulated for 13 weeks of hourly system load data (actually net dispatchable system load data, as mentioned above), to sample the effects of resource addition over a range of demand conditions. The results in Table I are averages for the 13 load scenarios.

The net profits seen by a completely dispatchable 1000 MW resource average 2.98% of the base case estimated operating costs. Each of the simulated departures from complete dispatchability increases, or at any rate does not decrease, the constraints on system operations. Each constraint on dispatchability must therefore increase the costs of operating the system at maximum efficiency. The costs of the imposed constraints are lowest for the 16 hour minimum down time constraint, whose cost averages 0.19% of base case operating costs. Most of the profits from dispatch under this constraint are realized during peak hours, and if the profits are high enough there is no restriction against continuing a dispatch overnight at the 100 MW minimum level.

In the case of a $50,000 fixed cost added to the $25/MWH variable costs, operating profits are reduced by an average of 0.53% of base case operating costs, approximately three times the loss due to the 16 hour minimum down time constraint. There is less incentive in this case than in the minimum down time case to "cycle" the resource, that is to shut the resource down during the off-peak hours, because an additional fixed cost is then incurred at the next startup. Keeping the resource on off-peak when variable costs are above payments reduces the resource's net profits.

The case of 8 hours maximum up time corresponds more closely to a dispatchability regimen likely to be acceptable to interruptible customers, who are unlikely to agree to the utility's continuing interruptions overnight, even at minimum levels, in order to avoid violating a minimum down time constraint or paying a fixed charge. As mentioned above, this case also includes a 16 hour minimum down time constraint, because without a minimum down time it is possible to simply take a series of 8 hour dispatches with no real break between them, without violating the maximum up

TABLE I:
Production Costs, Utility Profits and Profit Losses for Simulated Operations with an Incremental Resource (when it is owned by the utility)

Resource Characteristics	**Average Production Cost**	**Average Utility Profits**	**As % of Base Costs**	**Average Profit Loss**	**As % of Base Costs**
Base Case: No Resource	$10,968,356.9	-	-	-	-
Complete Dispatchability	$10,639,787.0	$328,569.9	2.98	-	-
16 Hours Min. Down Time	$10,660,163.1	$308,193.8	2.80	$ 20,376.1	0.19
$50K Startup Cost	$10,697,479.4	$270,877.5	2.45	$ 57,692.3	0.53
8 Hours Max. Up Time	$10,740,004.8	$228,352.1	2.10	$100,217.8	0.88
Maximum No. of Starts=3	$10,838,720.4	$129,636.5	1.20	$198,933.3	1.79

time constraint. In this case, the costs of the constraints average 0.88% of the base case operating costs. Thus almost half of the profits of the completely dispatchable resource have been lost, although the profits that remain still average 2.1% of the base case operating costs.

The most constrained case is that of a resource which may be dispatched at most three times over the week. This case includes an 8-hour maximum up time so that a single dispatch period does not continue overnight. There is, however, no minimum down time: three adjacent 8-hour dispatches, although obviously unlikely because of the low off-peak energy payments, would be permitted. These constraints reduce the total profits of the resource by about two-thirds: the remaining profits average 1.2% of base case costs, while the cost of the constraints averages 1.79% of base case costs.

VII. Simulation Results: Value to Utility vs. Resource Net Profits

In comparing integrated dispatch-based profits to post-dispatch prices, and the operations of an endogenously priced resource under the two pricing schemes, two measures are of interest. The first measure is utility payments to the resource; the second is net resource profits. Under nondiscriminatory pricing, total utility payments are $\sum_{t=1}^{T} \lambda_t p_{it}$ to the i'th resource over the scheduling horizon, where λ_t is the price offered per unit of energy produced in period t. Net resource profits are defined as above, as the difference between utility payments and resource operating costs.

If it is assumed that post-dispatch pricing is based on an optimal set of prices for the dispatch of the system without endogenously priced resources, the "supply-side" system, it is clear that these prices will tend to result in over commitment (since they should yield a close-to-feasible schedule for the supply-side system) when the new resources are introduced into the system. Therefore the payments and profits based on post-dispatch prices will in general serve as an approximate upper bound on the payments and profits under integrated pricing for any resource in the system, endogenously priced or not. On the other hand, integrated pricing can only reduce overall system operating costs. There is therefore likely to be a conflict between endogenously priced resources' preference for post-dispatch marginal cost pricing and the utility's preference for integrated dispatch marginal cost pricing, at least when the significance of endogenously priced resources in price-based system operations causes them to have a more than marginal impact on marginal costs.

Table II contrasts the net profits of the previously described endogenously priced resources under integrated dispatch pricing and under post-dispatch pricing. The net profits for the resources under a given set of prices are a byproduct of the Lagrangian relaxation algorithm described above. For integrated dispatch, the prices used are those from the near-optimal dual solution of the scheduling problem including the resource: these prices vary with the resource characteristics. For post-dispatch, the prices used are those from the near-optimal solution of the scheduling problem without the resource: these prices are independent of the resource characteristics.

It is clear that an incremental resource will always prefer post-dispatch to integrated dispatch prices. In the case of the completely dispatchable resource, average profits are $34,792.3 under integrated dispatch versus $325,029.1 under post-dispatch, an order-of-magnitude difference. The discrepancy is reduced when the operations of the resource are assumed to be constrained. For 16 hours minimum down time, average profits are $39,027.2 under integrated dispatch, $296,830.2 under post-dispatch. Given a $50,000 startup cost, profits average $10,607.3 under integrated dispatch and $264,140.6 under post-dispatch. For 8 hours maximum up time, profits average $82,717.5 under integrated dispatch and $195,969.3 under post-dispatch. And for a maximum of three dispatches over the week, average profits are $70,313.1 under integrated dispatch and $104,726.0 under post-dispatch.

Table II yields further information useful to both utility and resources as to the value of imposing operating constraints on resources. In the case of endogenously priced resources, such constraints may either be contractually defined or a matter of utility expectations. In either case, the constraints depend both on actual physical constraints on operations and on resource decisions about how to represent themselves to the utility.

Under post-dispatch pricing, the presence of operating constraints always reduces profits. This is evident, because the resource is maximizing the same net profit objective function with or without constraints. Hence constraints can only reduce the maximum profits. In the scenarios simulated, profit losses to the resource averaged $28,198.9 with 16 hours minimum down time, $129,059.8 with 8 hours maximum up time and 16 hours minimum down time, and $220,303.1 with a maximum of three dispatches and 8 hours maximum up time. The addition of a startup cost must also clearly reduce net profits, and in the simulations the profit reduction averages $60,888.5 (more than the startup cost itself). Thus, using post-dispatch pricing is an incentive

TABLE II
RESOURCE PROFITS UNDER INTEGRATED SCHEDULING VS. POST-DISPATCH SCHEDULING

Resource Charac.	Av. Proft. Integ. Sched.	Av. Proft. Post-dispatch	Difference
Complete Dispatchability	$34,792.3	$325,029.1	$290,236.8
16 Hours Min Down Time	$39,027.2	$296,830.2	$257,803.0
$50K Startup Cost	$10,607.3	$264,140.6	$253,533.3
8 Hours Max. Up Time	$82,717.5	$195,969.3	$113,251.7
Maximum No. of Starts=3	$70,313.1	$104,726.0	$ 34,412.9

for endogenously priced resources to behave, and therefore represent themselves, as nearly as possible as being completely dispatchable. Even when the post-dispatch policy may not appear to be the optimal pricing regimen for the utility to follow, these results indicate that the utility might for other reasons use such a policy when its benefits, such as increased reliability of endogenously priced resource operations, outweigh lost utility revenues. These potential benefits are not, however, captured in the short-term scheduling model, due to its deterministic nature.

In contrast to post-dispatch pricing, integrated dispatch pricing does not necessarily reduce resource net profits when constraints are placed on the resource's operations. There is an interaction between the operating constraints and the modified set of prices obtained from the scheduling model, so that although the constraints ensure that the resource will be operated less than it would be if completely dispatchable, they also discourage its operation by the utility in hours less profitable to the resource. Thus under 16 hours minimum down time resource net profits increased an average of $4,234.9, under 8 hours maximum up time net profits increased $47,925.2 on average, and under a maximum of three dispatches net profits increased an average of $35,520.8. The average profit increases do however conceal the fact that under particular scenarios profits may decrease. For example, in the simulated week of highest profits for the completely dispatchable resource, profits decreased by $8,377.8 under 16 hours minimum down time and $33,600 under a maximum of three dispatches.

VIII. Use of the Short-Term Scheduling Model by Resource Operators

Because resource profits can increase under integrated dispatch in the presence of some operating constraints, there is some incentive for operators of resources to use such constraints as negotiating points in arranging a price-based operations policy with the price-setting utility. At worst, operators of resources might find it advantageous to lie about their operating characteristics by placing many unnecessary constraints on their operations by the utility in order to maximize their profits. A separate side payment, not figuring in a resource's short-term response to prices, might be used as an incentive to get maximum dispatchability out of endogenously priced resources. This side payment could compensate resources for foregone profits without eliminating more significant cost benefits to the utility. It does appear, however, that marginal cost-based priced alone, if these marginal prices include accounting for a resource's own operations (as prices based on integrated dispatch do), are not adequate incentive for the resource to fully reveal its true operating characteristics.

It is, in fact, possible for operators of individual resources to use the utility's short-term scheduling model, if the solution procedures and system modeling assumptions are known, to determine the best level of constraint to offer to the utility operator. Moreover, the best level from the resource's perspective will not necessarily be the best level as seen by the utility, which clearly will want as few constraints as possible if removing constraints is without cost to the utility.

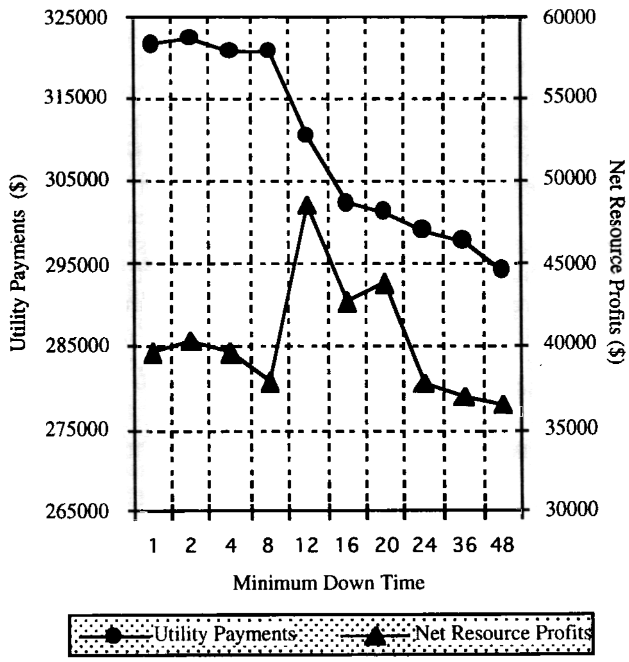

Figure 1: Utility Payments and Net Resource Profits vs. Minimum Down Time

Figure 1 shows the levels of utility payments (interpreted here as marginal value of the resource to the utility) and net resource profits, under minimum down times ranging

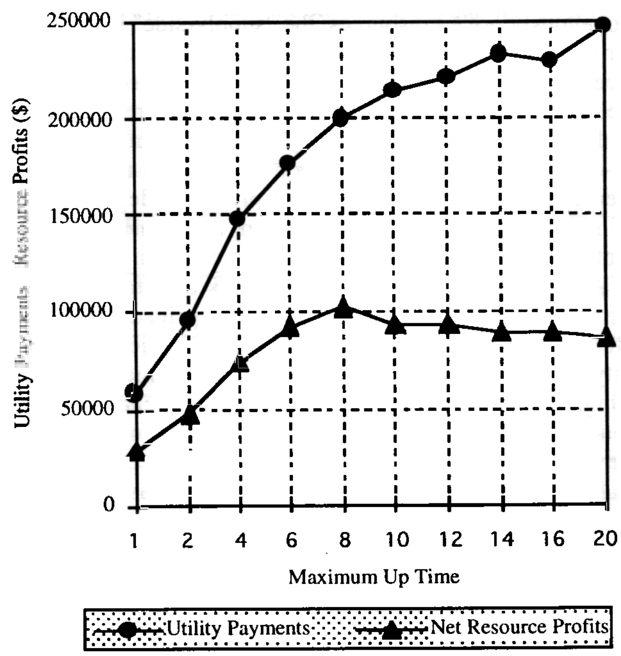

Figure 2: Utility Payments and Net Resource Profits vs. Max Up Time

between 1 and 48 hours, for a single week's simulated operations. From the utility's perspective, estimated marginal value of the resource is maximized at 2 hours minimum down time (1 hour minimum down time here shows lower value, presumably, because of "noise" in the solution procedure, since its true optimal solution should in theory show a higher value). However, from the resource operator's perspective, net profits are maximized at a minimum down time of 12 hours. If the resource's minimum down time is too long, the utility can at its discretion keep it up uneconomically over the off-peak hours, reducing resource profits.

Figure 2, above, shows estimated value and net profits, under 16 hours minimum down time and maximum up times ranging between 1 and 20 hours. The longer the maximum up time, as might be expected, the higher is the value derived by the utility. Net profits, however, are highest for a maximum up time of 8 hours, possibly because this combination of maximum up and minimum down times yields a natural repeatable daily cycle.

Finally, Figure 3 gives the values and resource net profits under a maximum up time of 8 hours and maximum number of allowed dispatches per week ranging between one and six. Value rises with the maximum number of allowed dispatches, but net profits are highest at five allowed dispatches. With six allowed dispatches, the optimal schedule includes two adjacent dispatches on one day, because the interruptions are much more valuable on weekdays than on weekend days. Hence at six dispatches some hours receive off-peak prices for generation, and resource net profits are reduced.

The above examples provide a methodology for the utility

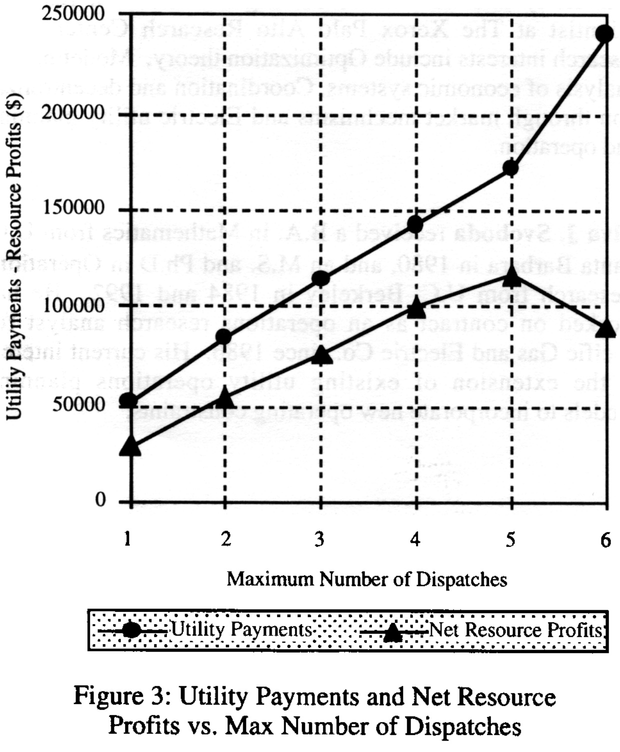

Figure 3: Utility Payments and Net Resource Profits vs. Max Number of Dispatches

and potential endogenously priced resources to determine, for negotiating purposes, the level of side payment required to make the customer indifferent between reporting its operating constraints truthfully and therefore being as dispatchable as possible, versus choosing an "untruthful" but profit-maximizing constraint level.

IX. Conclusions

We have shown that it is appropriate to use a near-optimal solution of the Lagrangian dual of the short-term scheduling problem as a simulation of system operations, when the infeasibilities of the corresponding resource schedule with respect to the demand requirements are of a magnitude comparable to the inherent error in the demand forecasts. It was also demonstrated that, for certain purposes, simulation of integrated operations including endogenously priced resources, is more appropriate for determining spot prices than the use of Lagrange multipliers from a purely supply-side (i.e., utility-owned and utility-controlled) simulation. In particular, benefits may accrue to the the utility from using integrated dispatch to estimate the optimal spot prices to be used in the dispatch of the endogenously priced resources.

The short-term scheduling problem's dual solution has a significant advantage for price-setting in the context of price-based system operations because it yields a set of prices that, when communicated to all resources (assuming they are modeled correctly) produces near-optimal, near-feasible schedules of operations of the power system over a given scheduling horizon. It is also suggested that for a system with the ability to "tune up" these schedules, either by using resources specifically dedicated to following load fluctuations, or by economy interchanges with neighboring utilities, or both, the integrated scheduling model provides a realistic basis for price-based operations and the setting of spot prices.

However, if the resource models used in integrated scheduling include operating characteristics similar to those that constrain supply-side resources at present, endogenously priced resources may maximize their expected net profits by deliberately misrepresenting their operating characteristics to the utility. If such misrepresentation is viewed by the utility as a bad thing, it must be remedied outside the operational context. Examples of such remedies might include bidding weights based on the stated operating characteristics, or charging for the constraints contractually.

Nonetheless, we have concluded that the complexity of the constrained short-term resource scheduling process is not an impediment to practicable price-based power system operations, or to the incorporation of endogenously priced resources and programs, such as real-time pricing or priority service [3], into these operations. Indeed, use of integrated scheduling allows the benefits of endogenously priced resources to be more clearly seen by the utility in the process of designing pricing schemes and contracts that will be attractive both to itself and to the resources who understand their marginal valuation and cost characteristics well enough to want to participate in price-based operations in the first place.

As a final note, we should mention that, as pointed out by one of the referees, our proposed methodology addresses, directly, only constraints on price-based resources that are defined within the time frame of short term scheduling. This excludes, for instance, some types of constraints addressed in [2] such as an annual limit on the number of interruptions (start-ups of the price-based resource). One possible option for dealing with such constraints is to augment the operating cost of the resource with a shadow price reflecting the long term constraint (imported from a long term scheduling program). Similar remedies have been used in short term scheduling of hydro resources by placing a value on water to reflect long term constraints.

X. ACKNOWLEDMENTS

The first author acknowledges the invalulable personal assistance of Ali Vojdani and Fulin Zhuang in his doctoral research which underlies this paper. We also thank Pacific Gas and Electric Co. for allowing the use of a simplified version of its Hydro-Thermal Optimization program in our numerical experiments. Todd Strauss was extremely helpful in providing us the CALECO system data for the simulations. This work was partially supported by a grant from the University of California Universitywide Energy Research Group. The second author has also been supported by NSF grant IRI-8902813

XI. REFERENCES

[1] D.P.Bertsekas et al, "Optimal Short-term Scheduling of Large-scale Power Systems," *IEEE Trans. Automatic Control*, vol. AC-28 no. 1, 1983, pp. 1-11.

[2] D.W. Caves and J.A. Herriges, "Optimal Dispatch of Interruptible and Curtailable Service Options," *Operations Research* vol. 40, 1992, pp. 104-112.

[3] Hung-po Chao and R.B. Wilson, "Priority Service: Pricing, Investment, and Market Organization," *American Economic Review*, vol. 77 no. 5, 1987, pp. 899-916.

[4] Decision Focus, Inc., "The DYNAMICS Model for Measuring Dynamic Operating Benefits", EPRI GS-6401 Project 2999-6 Interim Report, 1989.

[5] L.A.F.M. Ferreira, T. Andersson, C.F. Imparato, T.E. Miller, C.K. Pang, A. Svoboda and A.F. Vojdani, "Short-term Resource Scheduling in Multi-area Hydrothermal Power Systems," *Electrical Power & Energy Systems*, 1989.

[6] E. Kahn, *Electric Utility Planning and Regulation*, American Council for an Energy Efficient Economy, 1988.

[7] C. Marnay and T. Strauss, "Chronological Model Comparison," Report to the Calif. Public Utilities Commission, 1990.

[8] A. Merlin and P. Sandrin, "A New Method for Unit Commitment at Electicite de France," IEEE/PES Summer Meeting, 1982.

[9] J.A. Muckstadt and S.A. Koenig, "An Application of Lagrangian Relaxation to Scheduling in Power-Generation Systems," *Operations Research* vol. 25, 1977, pp. 387-403.

[10] B.T. Polyak, "Minimization of Unsmooth Functionals," *U.S.S.R. Computational Math. Math. Physics* vol. 9, 1969, pp. 14-29.

[11] F. Zhuang, *Optimal Generation Units Commitment in Thermal Electric Power System,* Ph.D. Thesis, Department of Electrical Engineering, McGill University, Montreal, Canada, December 1988.

Shmuel S. Oren is Professor of Industrial Engineering and Operations Research at the University of California at Berkeley. He holds a B.S and M.S in Mechanical Engineering from the Technion in Israel and an M.S. and Ph.D. in Engineering Economic Systems from Stanford. Prior to his current position he was on the faculty of The Engineering Economic Systems Department at Stanford University and he worked for eight years as a research scientist at The Xerox Palo Alto Research Center . His research interests include Optimization theory, Modeling and analysis of economic systems, Coordination and decentralization through market mechnisms and Electric utility planning and operation.

Alva J. Svoboda received a B.A. in Mathematics from U.C. Santa Barbara in 1980, and an M.S. and Ph.D in Operations Research from U.C. Berkeley in 1984 and 1992. He has worked on contract as an operations research analyst for Pacific Gas and Electric Co. since 1986. His current interest is the extension of existing utility operations planning models to incorporate new operating constraints.