J. Webster (ed.), *Wiley Encyclopedia of Electrical and Electronics Engineering*

HYDROTHERMAL POWER SYSTEMS

A modern power system usually contains various types of generating sources including thermal (nuclear, fossil steam, gas turbine), hydro (run-of-the-river and regular plants), and pumped-storage units, as well as other renewable sources such as wind and solar. The primary goal of hydrothermal power system operation is to provide reliable and economical power (1) for electricity users. To realize this goal, many operational issues have to be addressed, including those at a snapshot such as automatic generation control and optimal power flow (see Power system control) and those over a time horizon. One of the most important daily tasks is hydrothermal scheduling or hydrothermal coordination. This determines when and at what levels units should be run to meet the system demand and reserve requirements while satisfying individual unit constraints with a minimum total generation cost. The economic consequence of operation scheduling is significant. A small percentage of reduction in generation costs may result in savings of millions of US dollars for a large utility company (2). With deregulation of the electric utility industry a global trend, the daily scheduling activities may be changed for the new environment. The task of coordinating generating resources to minimize costs or to maximize profits, however, will become even more important.

The operating characteristics and generation costs of generators differ significantly from one unit to another. Identifying these characteristics and properly coordinating their operations can make full use of their advantages to reduce total generation cost. Among thermal units, the operating costs of nuclear units are cheaper than others. These units are thus often used as "base load units" with generation levels fixed at the full capacity in view of economic, safety, and environmental reasons. Most large fossil steam units are also base load units, whereas others operate in cycles following the system demand with additional costs and constraints associated with their startups and shutdowns. Gas turbines can be quickly started but have high fuel costs and are often used to satisfy peak demand. Compared to the costs of thermal units, generation costs of hydro units are usually insignificant. However, their "fuel," the natural water inflow, is limited within a time frame, and their operations may be coupled because of the hydraulic connection of reservoirs within a river catchment. A pumped-storage unit pumps water into its upper reservoir during the period of low demand, using the economical energy generated by base load units, and generates at the time of high demand and high costs. It can smooth peak load, provide reserve, and play an important role in reducing the overall generating costs.

Based on the goals of operations and the time horizon under consideration, hydrothermal scheduling can be classified into several levels of hierarchy as shown in Fig. 1.

Nuclear units need to be refueled once every few years, and the plants must be shut down several weeks during refueling for maintenance. Preventive maintenance is also needed for other units and is usually scheduled over a horizon of one to several years. Midterm hydro scheduling or seasonal hydro planning produces hydro schedules over a time frame from several months to a year, and determines how to distribute the total water resource among the weeks under consideration. In the process, it is necessary to estimate natural water inflows, system demand, and unit unavailability due to unscheduled unit outages based on historical and forecast data. Maintenance scheduling in conjunction with midterm hydro scheduling ensures that units and hydro energy are available at the times of high demand. Taking weekly hydro energy allocated by the midterm scheduling as an input, the short-term hydrothermal scheduling generates schedules for thermal and hydro

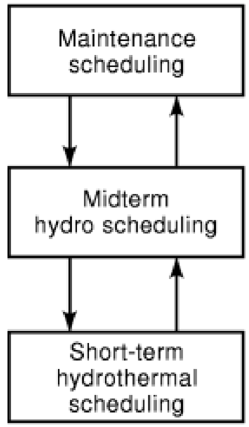

Fig. 1. Hierarchies of hydrothermal scheduling.

units on an hourly basis over a horizon from one day to ten days. It encompasses two interrelated problems: (1) the simultaneous determination of the "unit commitment" (unit up and down) status and preliminary hourly generation levels for all units and (2) the "economic dispatch" of power (i.e., the determination of hourly generation levels) given the unit commitment status. The second problem is a special case of the first. Although hydrothermal scheduling provides a preliminary power output for each unit, actual generation levels are decided by an economic dispatch calculation on an hourly basis using the latest demand and cost information. The focus of this article is on short-term hydrothermal scheduling. The approaches developed for short-term scheduling, however, can often be used for midterm and longterm scheduling problems.

Problem Description

Short-term hydrothermal scheduling has to consider many factors including

- unit generation costs
- system-wide constraints
- individual unit operating constraints

Generation Costs. Generation costs typically include fuel (running) costs, startup costs, and maintenance costs. Generally only fuel and startup costs of thermal units are considered in the short-term hydrothermal scheduling. The fuel cost of a thermal unit depends on its heat-rate characteristics and can be expressed as a load-dependent incremental cost plus an overhead called no-load cost existing all the time as long as the unit is running—resulting from the heat loss, wear and tear, etc. This no-load cost is different from capital charge, which is incurred whether the unit is running or not and is irrelevant to scheduling. Thus if a unit is shut down, its generation cost is assumed to be zero. Typically when a unit is running, its generation cost can be expressed by a polynomial function:

$$C[p(t)] = C_0 + \sum_{m=1}^{n} k_m p^m(t) \qquad (1)$$

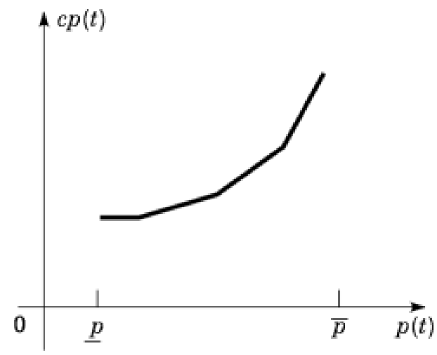

Fig. 2. Piece-wise linear cost function.

where $p(t)$ is unit power output at time t, and C_0 the no-load cost. For many units, $C[p(t)]$ can be well approximated by a quadratic, piecewise quadratic, or by an n-segment piecewise linear and convex function:

$$C[p(t)] = \max\{C_0 + k_1 p(t),\ C_1 + k_2 p(t), \ldots, C_{n-1} + k_n p(t)\} \tag{2}$$

as shown in Fig. 2, where $\tilde{p}$ is the generation capacity and is the minimum generation. The cost curve in Eq. (1) is smooth, and the cost in Eq. (2) is computationally convenient for economic dispatch. The forms of the cost functions, however, are not essential for solving hydrothermal scheduling problems.

The startup cost of a thermal unit is made up of the fuel and operating costs to bring the unit on-line; it is a time-varying functuion because it depends on the unit status (i.e., the time since last shutdown). In the literature, this cost is often assumed to be an exponential function of unit down time as

$$S(t) = C_s\left(1 - e^{-t/d}\right) \tag{3}$$

where t is the time since last shutdown, C_s a positive constant representing the cold startup cost, and d is the boiler cool-down coefficient. For certain units, the startup cost is modeled as a linear function of unit downtime, while remaining constant after the cold startup time (3). The startup or shutdown costs for a hydro or a pumped-storage unit are usually ignored because they are insignificant as compared to those of a thermal unit. Frequent startups and shutdowns, however, may increase maintenance costs (4). Therefore, penalty terms can be added to discourage excessive startups and shutdowns of hydro and pumped-storage units.

System-Wide Constraints. The system demand or load balance equation is written as

$$\sum_i p_{ti}(t) + \sum_j p_{hj}(t) + \sum_k p_{pk}(t) = P_d(t) \tag{4}$$

where $p_{ti}(t)$, $p_{hj}(t)$, $p_{pk}(t)$ are generation levels of thermal unit i, hydro unit j, and pumped-storage unit k at time t, respectively; $P_d(t)$ the total system demand at time t containing the aggregated customer demand and losses (approximately accounted for by solving an optimal power flow problem). The generation of a pumped-storage unit may take negative values representing pumping. If transactions among utilities are considered, the system demand in Eq. (4) should also add the net power sold and deduce the net power purchased.

The system reserve requirements include spinning and supplemental reserves which are the extra power that the system should be able to provide within specified times guard against disturbances such as unscheduled outage of a unit or unforeseen load changes. The spinning reserve is the reserve that on-line units can provide. Generally, the spinning reserve contribution of an on-line unit is the amount of additional power that the unit can provide within, say, 10 min {i.e., $r[p(t)] = \min[\tilde{p} - p(t), \tilde{r}]$, where $p(t)$ is the unit's power output, $\tilde{p}$ is the generating capacity, and $\tilde{r}$ is the maximum reserve contribution determined by the unit's ramping capability}. The reserve is zero if the unit is down, except for some specifically designated gas turbines that can be started very quickly. The system-wide spinning reserve constraints require that the total reserve contribution should be no less than a specified requirement $P_r(t)$ at any time; in other words,

$$\sum_i r_{ti}(t) + \sum_j r_{hj}(t) + \sum_k r_{pk}(t) \geq P_r(t) \tag{5}$$

Supplemental reserves such as 10 min nonspinning reserve or 30 min reserve can be defined in the similar way.

In short-term scheduling, the capacity of transmission networks is often assumed to be large enough, and transmission constraints are generally ignored. With the deregulation of the electric power industry and open access of transmission networks, there is an increasing pressure not only to optimize the use of generating resources but also to utilize transmission networks optimally (see Electricity supply industry). Therefore the capacity of transmission networks, viewed as power flow limits on branches and/or voltage bounds on buses, is becoming important in operation scheduling. Integrated resolution of the scheduling and transmission problem, however, is complicated and difficult (see Electricity supply industry and Electricity supply industry) in view of the sizes of practical transmission networks. Adding a large number of nonlinear transmission constraints in hydrothermal scheduling would drastically increase computational requirements. Therefore, transmission networks are usually simplified to a dc power flow model (5,6), or transmission constraints are indirectly considered by limiting the generations of individual regions (7). Nevertheless, methods for solving integrated scheduling and optimal power flow problems have been presented with limited success (8,9).

Individual Unit Operating Constraints. Individual unit constraints may be quite different from one unit to another in view of the differences in unit characteristics and principles of operation. A nuclear unit seldom changes its generation level unless absolutely necessary because of economic and safety considerations. Run-of-the-river hydro units are not scheduable (i.e., the natural water inflows are used instantaneously for generation). Almost all units are subject to minimum generation and capacity constraints as in Eq. (6)

$$\begin{aligned} &\underline{p} \leq p(t) \leq \overline{p} && \text{if unit is up} \\ &p(t) = 0 && \text{if unit is down} \end{aligned} \tag{6}$$

The minimum up (down) time constraints specify how long a unit must be kept on (down) before it can be shut down (started up). They are imposed on large thermal units because of heat stress on the equipment, and on hydro units to prevent frequent startups and shutdowns. Ramp rate constraints specify a unit's maximum allowable generation change between two consecutive time instances, as described by

$$p(t-1) - \Delta \leq p(t) \leq p(t-1) + \Delta \tag{7}$$

where Δ is the unit's ramp rate. The ramp rate constraints couple the unit's generation levels over the scheduling horizon and are difficult to deal with. In addition, some units can be "must-run," or "must-not-run," or must generate at predetermined levels during specified periods for scheduled maintenance or testing purposes.

The power output of a hydro unit is a function of the amount of water discharge and reservoir head—the elevation difference between the upstream and downstream water surfaces of a reservoir (see Hydroelectric

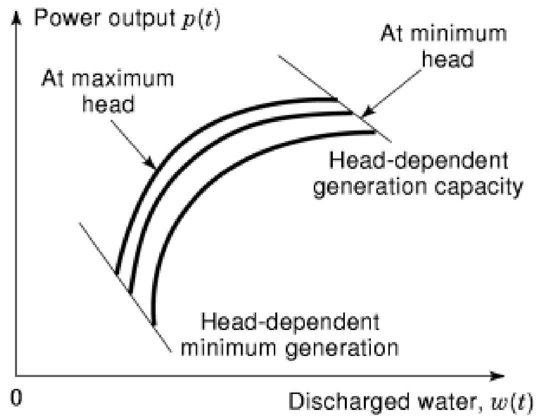

Fig. 3. A typical hydro power output versus discharged water and reservoir head.

power stations). A typical power output versus discharged water and head is shown in Fig. 3. The minimum and maximum water discharge, and thus the minimum and maximum power output, can also be head-dependent. Because the reservoir head is a function of the volume of water stored given the shape of the reservoir bed, the power output is thus generally modeled as a function of the volume of water stored and water discharge. Forbidden operating regions are common for many units (i.e., a unit may not operate in certain regions or may even be limited to operate at discrete points for safety or efficiency reasons). Pumped-storage units have characteristics similar to those of hydro units (see Hydroelectric power stations) and can significantly contribute to system reserve in both generating and pumping modes.

Scheduling hydro units is generally more difficult than scheduling thermal units because the reservoirs in a river catchment or water shed are hydraulically coupled. The water discharged by an upstream unit affects the levels of the downstream reservoirs. The water flow in a river catchment can be described by the following the water balance equation

$$\begin{aligned} v_j(t+1) = {} & v_j(t) - [w_j(t) + s_j(t)] \\ & + \sum_{i \in U_j} [w_i(t-\tau_i) + s_i(t-\tau_i)] + \zeta_j(t) \end{aligned} \tag{8}$$

where $v_j(t)$, $w_j(t)$, and $s_j(t)$ are, respectively, the reservoir content, water discharge, and spillage of hydro unit j at time t, $\zeta_j(t)$ is the sum of natural inflows to reservoir j at t, τ_j is the time required for water discharged from j to reach its directly downstream reservoir, and U_j is the set of direct upstream reservoirs of reservoir j. Therefore the schedules of hydro units in a river catchment are coupled not only through system-wide demand and reserve requirements in Eqs. (4) and (5) but also the water balance in Eq. (8). The hydro scheduling problem can also be viewed as a resource allocation problem—there is limited hydro energy in a cascaded reservoir system, and the problem is to determine how to allocate the energy to maximize the benefits or savings.

The water volume in a reservoir is also constrained by the reservoir capacity and its minimum (or dead) volume as

$$\underline{v}_j \leq v_j \leq \overline{v}_j \tag{9}$$

From this analysis, the short-term hydrothermal scheduling involves discrete decision variables such as unit on/off and continuous decision variables such generation levels. It can thus be formulated as a mixed-integer programming problem:

$$\min_{p_{ti}(t), w_{hj}(t), w_{pk}(t)} C, \text{ with } C = \sum_t \left\{ \sum_i [C_{ti}(p_{ti}(t)) + S_{ti}(t)] + \sum_j S_{hj}(t) + \sum_k S_{pk}(t) \right\} \quad (10)$$

where $S_{ti}(t)$, $S_{hj}(t)$, and $S_{pk}(t)$ are, respectively, the startup costs for thermal, hydro, and pumped-storage units, subject to system-wide generation, reserve, and possibly transmission constraints, water flow balance constraints of each river catchment, and individual unit constraints. This problem mathematically belongs to the class of *NP* (Nonpolynomial) hard combinatorial optimization problems (i.e., the computational time required to find an optimal solution escalates exponentially with problem size). Consistently obtaining optimal schedules has been proven to be extremely difficult for systems of practical sizes. Nevertheless, because the cost to be minimized is the sum of individual unit costs and the system-wide demand and reserve requirements are also unit-wise additive, the problem is "separable," and efficient near-optimal approaches can be developed as will be presented later.

For midterm and longterm scheduling problems, since many parameters are stochastic in nature, uncertainties may have to be considered. Examples include demand, unit availability, and natural inflows of reservoirs. For instance, if a possible flood is not appropriately considered, hydro energy may be forced to spill out during the flood. On other hand, if an unusual draught is not properly taken into account, the hydro energy may not be used when it is most needed. By dividing a year into several seasons, it is possible to approximate the demand and inflows as stationary series using Markov chains, and the problem is to minimize the expected value of the cost function in Eq. (10). The resulting solution is a "decision policy" detailing what to do under each possible circumstance, as opposed to a sequence of decisions for a deterministic problem. The decision policy can, in turn, provide targets or constraints for the short-term scheduling problem.

Solution Methodologies

The approaches for solving hydrothermal scheduling problem can be classified into five categories (2):

- Partial enumeration such as branch and bound,
- Dynamic programming,
- Bender's partitioning,
- Heuristics,
- Near-optimal methods such as Lagrangian relaxation and its extensions.

Branch-and-bound is a tree search method with various branching and bounding (pruning) techniques carried on a tree as in Fig. 4.

A parent node within the tree represents a problem simpler than the original one with some constraints (typically integrality requirements on some discrete variables) relaxed and some variables fixed at specified values. The child nodes are a set of problems with more variables specified. The cost of the optimal schedule represented by a parent node is therefore a lower bound on those for its child nodes. If the cost of any feasible schedule is lower than the optimal cost of a node, the schedules corresponding its children should be discarded. This "bounding process" may significantly reduce the space to be searched. The three elements of the branch-and-bound method are thus the representation of a problem in terms of a search tree, the method for solving problems associated with each node, and the method for searching (branching and bounding) the

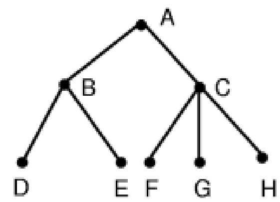

Fig. 4. Search tree of the branch-and-bound method.

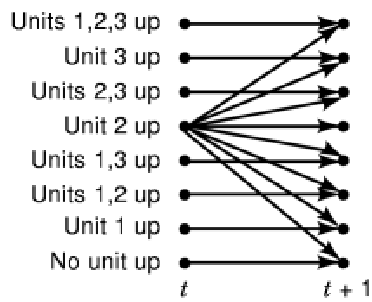

Fig. 5. The state transition diagram. For clarity, only the transitions from the state "unit 2 up" are shown, and transitions from other states are similar to those from state "unit 2 up."

tree. Many methods can be used to solve problems associated with individual nodes. One such method is to allow all 0–1 commitment decision variables to take in real values between 0 and 1 for the problem associated with the root node. Then, for each child of a node, gradually set certain commitment variables to either 0 or 1. The optimization problem at each node is thus reduced to a standard linear or nonlinear programming problem that can be efficiently solved by using an appropriate continuous variable optimization algorithm. Many approximations and modifications have been developed to improve the efficiency of this method.

Dynamic programming based on backward (or forward) searching of a state space has been extensively used in many areas. In its applications to hydrothermal scheduling, stages correspond to hours, and states are associated with unit commitment status. Figure 5 illustrates the states of two consecutive stages for a simple system with 3 thermal units without considering their minimum up and down times. Calculating the optimal fuel cost $C[p_{ti}(t)]$ for a given commitment status requires solving an economic dispatch problem, typically a constrained nonlinear programming problem, to produce the generation levels of all units. The optimal "cost-to-go" functions are then established for all possible states by working backward (or forward) along the scheduling horizon.

The optimal decision at each state is then obtained by traversing from the initial (or terminal) state. It can be seen that the number of states at each time unit for this simple configuration is 2^I, a very large number if number of units I is large. When minimum up and down times come into the picture, a single up state as in Fig. 5 cannot fully represent the status of a unit because the unit may not be able to shut down within the next few hours. The number of states required to describe the problem will thus be increased to M^I, where M is the sum of minimum up and down times. Dynamic programming can also be used to include hydro and pumped-storage units. Many constraints such as discontinuous operating regions and even discrete generation levels can be incorporated. However, because the continuous reservoir levels have to be discretized, a very large state space is required.

The idea of Bender's partitioning or Bender's decomposition is to decompose a problem into a "master problem" involving "complicating" (usually discrete) variables and a "slave problem" involving other (usually continuous) variables (2). The slave problem typically corresponds to economic dispatch with fixed unit commitment status. The marginal costs obtained from the slave problem are used to limit the commitment of the units in the master problem, which can be further decomposed into a set of subproblems associated with individual units. The master problem then provides a new set of commitment decisions. The two problems are iteratively solved until the solutions converge.

For branch-and-bound, dynamic programming, and Bender's decomposition, the "curse of dimensionality" is apparent when there are many units in the system, and these methods can hardly be used to schedule problems of practical sizes. Approximations and heuristics have to be applied at the sacrifice of optimality. One typical method to simplify the problem is to decompose it into two subproblems: hydro subproblem and thermal. The marginal costs obtained from thermal scheduling are used as a basis to allocate hydro energy over the hours. Then in thermal scheduling, hydro contributions are deducted from the system demand and reserve requirement. Hydrothermal coordination is realized by iteratively updating the hydro and thermal solutions. This iterative coordination scheme is heuristic in nature, thus optimality and sometimes convergence cannot be guaranteed.

Heuristic methods are still widely used by many electric utilities because of their simplicity and computational efficiency. These methods are often developed based on the knowledge of experienced operators for specific systems. One method is the so-called priority list commitment. Thermal units are first ranked based on their full-load average costs, and units are committed in the ascending order of the cost to meet system demand and reserve requirements. Hydro energy is then allocated to "shave" the peak of system demand (having highest incremental costs). Manual adjustment is needed to satisfy individual unit constraints. The disadvantages of heuristics are that the methods are ad hoc for specific systems, manual adjustment can be very time consuming, and the schedule obtained may be far from optimal.

Lagrangian relaxation and its extensions have consistently been proven to provide near-optimal solutions in a computationally efficient manner and can be easily adapted to different problem configurations (3,5,6,7,8, 10,11,12). They are thus among the most successful and most widely used methods and deserve more attention here.

Lagrangian relaxation has been fundamental for deriving *necessary and sufficient conditions for optimality.* Over the decades, it has been the corner stone of resource allocation theory in economics. In mathematics, it has also shed much light on the development of efficient algorithms for constrained optimization. The method is particularly powerful for combinatorial optimization of "separable problems" where the objective function and system-wide *coupling constraints* are *additive* in terms of basic decision variables. For the scheduling problem under consideration, because the total cost to be minimized is the sum of individual unit costs, and the demand and reserve requirements are also unit/transaction-wise additive, the problem is "separable," and Lagrangian relaxation can be effectively applied.

The key idea of the approach, in a nut shell, is *decomposition* and *coordination.* "Hard" system demand and reserve requirements are first "softened" or "relaxed" by having to pay a "price" or "penalty" for constraint violations. The per unit violation penalty at a particular hour is the hour's "Lagrange multipliers," or the "shadow prices" in the economics literature. Because the original problem is separable, the "relaxed problem" can be decomposed into many smaller subproblems, one for each unit or transaction. Given a set of multipliers, each unit determines its generation levels across the time horizon to minimize its cost and subject to its own constraints, and similarly for transactions. These subproblems are not NP-hard, and can be efficiently solved by using "dynamic programming." The multipliers are then iteratively adjusted based on the levels of constraint violation following the market economy concept (i.e., increase the multipliers for undergenerated hours and reduce the multipliers for overgenerated hours). At the termination of such multiplier-updating iterations, system-wide constraints may still be violated at a few hours. Simple heuristics are then applied to adjust subproblem solutions to form a feasible schedule satisfying all constraints. Because the value of the dual problem is a lower bound of the optimal cost, the quality of the feasible schedule can be quantitatively evaluated.

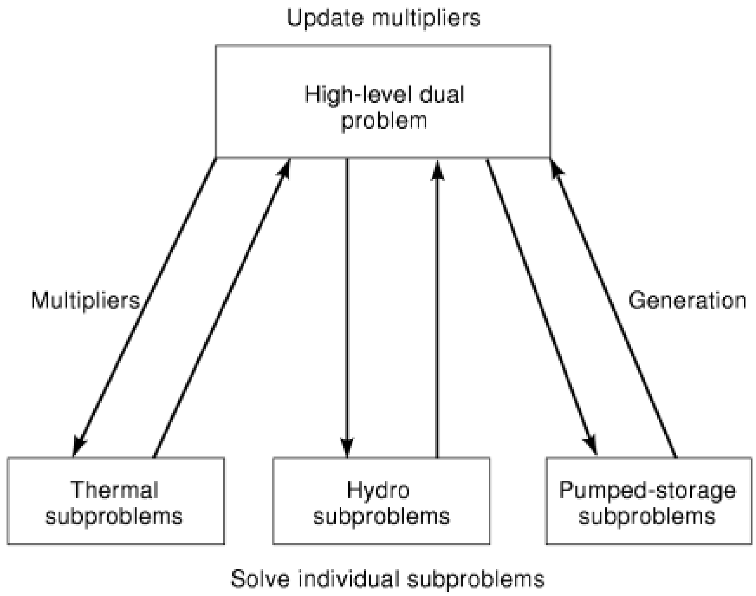

Fig. 6. Framework of the Lagrangian relaxation method.

The resolution of the original problem is thus performed through a two-level approach as shown in Fig. 6, where the *low level* consists of solving individual subproblems. Coordination of subproblem soluotions is done through the iterative updating of Lagrange multipliers at the *high level* to ensure near optimality of the overall solution. In optimization terminology, the *nondifferentiable concave* "dual function is *iteratively maximized* during the multiplier updating process.

To be more specific, the "Lagrangian" is first formulated as follows:

$$
\begin{aligned}
L = \sum_t \Bigg\{ & \sum_i [C_{\mathrm{t}i}(p_{\mathrm{t}i}(t)) + S_{\mathrm{t}i}(t)] + \sum_j S_{\mathrm{h}j}(t) + \sum_k S_{\mathrm{p}k}(t) \\
& + \lambda(t)\left[P_{\mathrm{d}}(t) - \left(\sum_i p_{\mathrm{t}i}(t) + \sum_j p_{\mathrm{h}j}(t) + \sum_k p_{\mathrm{p}k}(t)\right)\right] \\
& + \mu(t)\left[P_{\mathrm{r}}(t) - \left(\sum_i r_{\mathrm{t}i}(t) + \sum_j r_{\mathrm{h}j}(t) + \sum_k r_{\mathrm{p}k}(t)\right)\right]\Bigg\}
\end{aligned} \tag{11}
$$

where $\lambda(t)$ and $\mu(t)$ are, respectively, the "Lagrange multipliers" associated with system demand and reserve requirements at time t. By using the duality theory (15) and the decomposable structure of Eq. (11), a two-level maximum–minimum optimization problem can be formed. Given the multipliers, the low level consists of the following subproblems:

Thermal subproblems ($\mathbf{P_t} - i$), $i = 1, 2, \ldots, \boldsymbol{I}$: min $L_{\mathrm{t}i}$, with

$$
L_{\mathrm{t}i} = \sum_t [C_{\mathrm{t}i}(p_{\mathrm{t}i}(t)) + S_{\mathrm{t}i}(t) - \lambda(t)p_{\mathrm{t}i}(t) - \mu(t)r_{\mathrm{t}i}(t)] \tag{12}
$$

subject to individual thermal constraints.

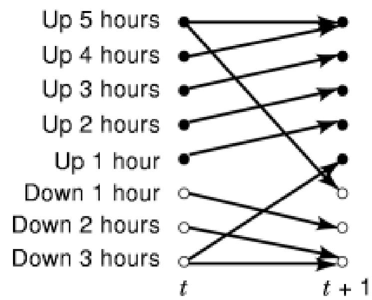

Fig. 7. The state transition diagram for a thermal subproblem. The minimum down time is assumed to be 3 hours, and the minimum up time, 5 hours.

Hydro subproblems ($\mathbf{P_h} - j$), $j = 1, 2, \ldots, \boldsymbol{J}$: min $L_{\mathrm{h}j}$, with

$$L_{\mathrm{h}j} = \sum_t [S_{\mathrm{h}j}(t) - \lambda(t)p_{\mathrm{h}j}(t) - \mu(t)r_{\mathrm{h}j}(t)] \tag{13}$$

subject to individual hydro constraints.

Pumped-storage subproblems ($\mathbf{P_p} - k$), $k = 1, 2, \ldots, \boldsymbol{K}$: min $L_{\mathrm{p}k}$, with

$$L_{\mathrm{p}k} = \sum_t [S_{\mathrm{p}k}(t) - \lambda(t)p_{\mathrm{p}k}(t) - \mu(t)r_{\mathrm{p}k}(t)] \tag{14}$$

subject to individual pumped-storage constraints.

Let $L^*{}_{\mathrm{t}i}[\lambda(t), \mu(t)]$, $L^*{}_{\mathrm{h}j}[\lambda(t), \mu(t)]$, and $L^*{}_{\mathrm{p}k}[\lambda(t), \mu(t)]$ denote the optimal sub-Lagrangians for ($\mathbf{P_t} - i$), ($\mathbf{P_h} - j$), and ($\mathbf{P_p} - k$), respectively. Then the high-level dual problem is max $\Phi[\lambda(t), \mu(t)]$, with

$$\begin{aligned}\Phi[\lambda(t), \mu(t)] = {} & \sum_i L^*_{\mathrm{t}i}[\lambda(t), \mu(t)] + \sum_j L^*_{\mathrm{h}j}[\lambda(t), \mu(t)] \\ & + \sum_k L^*_{\mathrm{p}k}[\lambda(t), \mu(t)] + \lambda(t)P_{\mathrm{d}}(t) + \mu(t)P_{\mathrm{r}}(t)\end{aligned} \tag{15}$$

subject to $\mu(t) \geq 0$. This is a two-level optimization framework. The solution process thus consists of (1) solving individual subproblems, (2) optimizing high-level multipliers, and (3) obtaining a near-optimal feasible schedule. The thermal subproblems, one for each unit, are typically solved by using dynamic programming (3). Stages correspond to hours within the scheduling horizon, and states within a stage correspond to the number of hours that the unit has been up or down as shown in Fig. 7, where the minimum up and down times are assumed as 5 and 3 hours, respectively. The optimal generation cost at a particular time instance for a particular state can be obtained by optimizing a single variable function, with startup and shutdown costs modeled as state transition costs. The optimal schedule of the unit for a given set of multipliers can then be obtained by using dynamic programming with a few states and well-structured state transitions. The "curse of dimensionality" caused by the coupling among units as mentioned earlier does not exist anymore because the units have been decoupled by using Lagrange multipliers.

Units with ramp rate constraints are harder to handle because generation levels of two consecutive hours are now coupled. Ramp rates can be handled through relaxation by introducing another set of multipliers to be updated at the intermediate level.

The hydro and pumped-storage subproblems are usually more difficult to solve because of the hydraulic coupling among the units within a river catchment. Network flow is an efficient algorithm for hydro subproblems. However, it cannot handle discontinuous operating regions (e.g., forbidden regions) nor discrete operating states (e.g., no generation). In this case, additional multipliers can be used to relax hydraulic coupling and updated at the intermediate level (16). Individual hydro subproblems can then be solved by dynamic programming similar to that for thermal subproblems, and a nonlinear network flow can be applied when the commitment status of hydro units is fixed.

The high-level dual problem is to update the multipliers so as to maximize the dual function in Eq. (15). Because discrete variables are involved at the low level, the dual function may not be differentiable at some points. The subgradient method is commonly used to update $\lambda(t)$ and $\mu(t)$ because of its simplicity and performance (9,17,18):

$$\lambda(t)^{l+1} = \lambda(t)^{l} + \alpha^{l} g_{\lambda}(t) \tag{16}$$

$$\mu(t)^{l+1} = \mu(t)^{l} + \alpha^{l} g_{\mu}(t) \tag{17}$$

where

$$g_{\lambda}(t) = P_{\mathrm{d}}(t) - \left[\sum_{i} p_{\mathrm{t}i}(t) + \sum_{j} p_{\mathrm{h}j}(t) + \sum_{k} p_{\mathrm{p}k}(t)\right] \tag{18}$$

is the subgradient of $\Phi[\lambda(t), \mu(t)]$ with respect to $\lambda(t)$, and

$$g_{\mu}(t) \equiv P_{\mathrm{r}}(t) - \left[\sum_{i} r_{\mathrm{t}i}(t) + \sum_{j} r_{\mathrm{h}j}(t) + \sum_{k} r_{\mathrm{p}k}(t)\right] \tag{19}$$

is the subgradient of $\Phi[\lambda(t), \mu(t)]$ with respect to $\mu(t)$, l is the high-level iteration index, and α is the step size which may be adaptively adjusted to speed up convergence (5,10). Other methods have recently been developed to update the multipliers with improved performance, including the reduced complexity bundle method (*RCBM*) (19). According to the duality theory, the multiplier $\lambda(t)$ is the shadow price or the marginal generation cost (i.e., the cost for generating an additional MW of power at time t). This is a by-product of Lagrangian relaxation and can be used to perform "what-if" analysis.

Subproblem solutions, when put together, are usually infeasible (i.e., the once relaxed system-wide constraints are generally not satisfied). Heuristic methods are thus used to modify subproblem solutions to form near-optimal feasible schedules. The generation levels of hydro, pumped-storage, and units with energy or ramp rate constraints are difficult to adjust because their generation levels are coupled across time through various unit-wise constraints. The schedules of these units are therefore first modified to satisfy their individual constraints if these constraints were violated because of the mid-level relaxation; then they remain fixed. The generation levels of the thermal units without these constraints are then adjusted to meet the system demand and reserve requirements based on subproblem solutions using heuristics (3,13,14).

An advantage of Lagrangian relaxation is that the dual cost $\Phi(t)$, $\mu(t)]$ defined in Eq. (11) is a lower bound to all feasible costs including the optimal one. Therefore the relative duality gap ε defined as

$$\epsilon = \{C - \Phi[\lambda(t), \mu(t)]\}/\Phi[\lambda(t), \mu(t)] \tag{20}$$

where C is the total generation cost of a feasible schedule as defined in Eq. (10) can be used to quantitatively measure the solution quality. If ε is small, the schedule is near optimal because the optimal cost $C*$ must lie between any feasible cost and the dual cost; in other words,

$$\Phi[\lambda(t), \mu(t)] \leq C^{*} \leq C \tag{21}$$

Another advantage of Lagrangian relaxation is its computational efficiency. Because the number of dual iterations generally does not increase as the number of units grows, computational requirements increase almost linearly as the problem size grows. A drawback of this approach is that a heuristic method is most likely needed to modify the dual solution to obtain a feasible schedule.

In summary, heuristic approaches such as priority-list have been widely used in practice because of their simplicity and insensitivity to modeling accuracy as compared to other approaches. Their disadvantages are obvious, and there is much room for improvement. At the early stage in developing optimization-based algorithms, rigorous approaches including branch-and-bound, dynamic programming, and Bender's partitioning were extensively investigated. These approaches, however, suffer from the "curse of dimensionality" and can hardly be put into practical use without approximation and simplification. Near-optimal methods such as Langrangian relaxation and its extensions have been developed for more than two decades and have become the main stream because of their near-optimal solution quality and computational efficiency. Because heuristics are involved to obtain feasible schedules, there is still room for improvement. Recent efforts include using evoloutionary computing approaches such as genetic algorithms to improve feasible schedules (20,21).

Solving scheduling problems with uncertainties usually requires stochastic optimization (17,18,22,23,24, 25). One of the approaches is stochastic dynamic programming. The idea is to extend dynamic programming to include states with probabilistic transitions and to use expected costs as the minimization objective. The direct consequence is that the state space and/or the number of possible transitions would increase significantly. For example, when stochastic dynamic programming is used to solve hydro scheduling problems with stochastic inflows, one more dimension will be added to the space to include probable inflows in addition to reservoir levels. To reduce computational complexity, reservoirs in a river catchment can be aggregated. The "aggregation–decomposition" method is to retain one reservoir and aggregate all the others into an equivalent one (18). The head effect on water-power conversion, however, can hardly be considered. Another method uses successive approximation in which reservoirs are solved one at a time by using stochastic dynamic programming while fixing the schedules of other reservoirs. The procedure is repeated until convergence is obtained; however, global optimality cannot be guaranteed. Lagrangian relaxation can also be applied where individual hydro subproblems with stochastic inflows are solved by using stochastic dynamic programming.

Another approach is optimization based on scenario analysis (17,25). The uncertain demand, unit availability, water inflows, etc., are modeled by a limited number of "scenarios" as shown in Fig. 8.

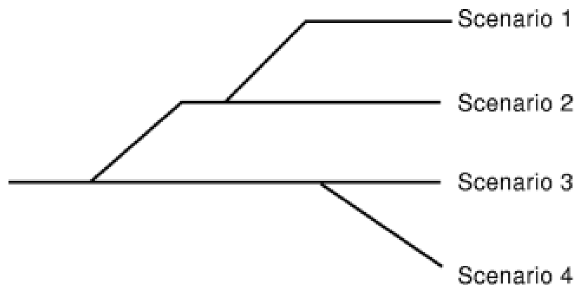

Fig. 8. Scenario tree.

Each scenario is associated with a probability of occurrence γ_m. The objective is to minimize the expected cost over M possible scenarios

$$\min_{p_{ti}^{\gamma_m}(t), w_{hj}^{\gamma_m}(t), w_{pk}^{\gamma_m}(t)} \sum_{m}^{M} \gamma_m \sum_{t} \left\{ \sum_{i} [C_{ti}(p_{ti}^{\gamma_m}(t)) + S_{ti}(t)] + \sum_{j} S_{hj}(t) + \sum_{k} S_{pk}(t) \right\} \quad (22)$$

Because the number of possible scenarios and consequently the computational requirements increase drastically as the number of uncertain factors increases, this approach can treat only a limited number of scenarios.

Conclusions

Hydrothermal scheduling is faced daily by electric power producers. It is a difficult mixed-integer programming problem with significant economic impact. Currently heuristic methods such as priority-list are still widely used in practice because of their simplicity. Near-optimal methods—Lagrangian relaxation and its extensions—seem to be most common among recent scheduling software packages.

Future challenges include the following two aspects. First, more and more constraints have to be considered. As mentioned in the section entitled "Problem Description," transmission constraints are being incorporated into scheduling (5,6,8). Other types of constraints might emerge or have emerged, for example, the environmental constraints limiting the amount of pollution from electricity generation (12). System parameters such as fuel prices and system demand are usually considered deterministic, but some of them may have to be modeled as stochastic or fuzzy in view of the uncertain nature of competitive markets. All these would complicate the scheduling problems. The challenge is to develop appropriate models and corresponding solution methodologies that capture the essence of the constraints while maintaining solution quality and computational requirements. Second, worldwide deregulation of the electric power industry would profoundly change the operations of hydrothermal systems therefore the associated scheduling problems. As presented earlier, a utility company is traditionally responsible for generating and delivering power to industrial, business, and residential customers in its service area. Under the new market structure, the utilities participate in electric markets. The resource scheduling problem still exists but has to be transformed and integrated with the market bidding structure. Market participants, including utilities and Independent Power Producers (*IPP*), have to decide how to submit bids and schedule generating resources to maximize their profit subject to various constraints, with bid selection determined by an Independent System Operator (*ISO*). The scheduling problem will not disappear, but will be transformed into a more complicated bidding and resource allocation problem. The results and insights obtained by researchers and practitioners over the decades on hydrothermal scheduling have established a solid foundation for us to tackle the new issues confronting the deregulated industry.

BIBLIOGRAPHY

1. A. Wood, B. Wollenberg, *Power Generation and Control*, Chichester, UK: Wiley, 1984.
2. A. Cohen, V. Sherkat, Optimization-based methods for operations scheduling, *Proc. IEEE*, **75**: 1574–1592, 1987.
3. X. Guan *et al.*, An optimization-based method for unit commitment, *Int. J. Electr. Power Energy Syst.*, **14** (1): 9–17, 1992.
4. O. Nilsson, D. Sjelvgren, Hydro unit start-up and their impact on the short term scheduling strategies of Swedish power producers, *IEEE Trans. Power Syst.*, **12**: 38–44, 1997.
5. J. J. Shaw, A Direct Method for Security-Constrained Unit Commitment, *IEEE Trans. Power Syst.*, **10**: 1329–1342, 1995.
6. S. J. Wang *et al.*, Short-term generation scheduling with transmission constraints using augmented Lagrangian relaxation, *IEEE Trans. Power Syst.*, **10**: 1294–1301, 1995.
7. A. F. M. Ferreira *et al.*, Short-term resource scheduling in multi-area hydrothermal power systems, *Elec. Power Energy Syst.*, **11** (3): 200–212, 1989.
8. R. Baldick, Generalized Unit Commitment, *IEEE Trans. Power Syst.*, **10**: 465–473, 1995.
9. H. P. Wolf, H. P. Crowder, Validation of subgradient optimization, *Math. Programming*, **6**: 62–88, 1974.
10. J. J. Shaw, D. P. Bertsekas, Optimal scheduling of large hydrothermal power systems, *IEEE Trans. Power Appar. Syst.*, **PAS-104**: 286–293, 1985.
11. A. Renaud, Daily generation management at Electricite de France: From planning towards real time, *IEEE Trans. Autom. Control*, **38**: 1080–1093, 1993.
12. El-Kaib, H. Ma, J. Hart, Environmetally constrained economic dispatch using Lagrangian relaxation method, *IEEE Trans. Power Syst.*, **9**: 1723–1729, 1994.
13. X. Guan *et al.*, Optimization-based scheduling of hydrothermal power systems with pumped-storage units, *IEEE Trans. Power Syst.*, **9**: 1023–1031, 1994.
14. X. Guan, P. B. Luh, L. Zhang, Nonlinear approximation method in Lagrangian relaxation-based algorithms for hydrothermal scheduling, *IEEE Trans. Power Syst.*, **10**: 772–778, 1995.
15. G. L. Nemhauser, L. A. Wolsey, *Integer and Combinatorial Optimization*, Chichester, UK: Wiley, 1988.
16. X. Guan *et al.*, An optimization-based scheduling algorithm for scheduling hydrothermal power systems with cascaded reservoirs and discrete hydro constraints, *IEEE Trans. Power Syst.*, **12**: 1775–1780, 1997.
17. P. Carpentier *et al.*, Stochastic optimization of unit commitment: A new decomposition framework, *IEEE Trans. Power Syst.*, **11**: 1067–1073, 1996.
18. A. Turgeon, Optimal operation of multireservoir power systems with stochastic inflows, *Water Resource Res.*, **16** (2): 274–283, 1980.
19. X. Guan, R. Baldick, E. Liu, Integrating power system scheduling and optimal power flow, *Proc. 1996 Power Syst. Comput. Conf.*, Dresden, Germany, 1996, pp. 717–723.
20. T. T. Mansfield, G. B. Sheble, Genetic-based unit commitment, *IEEE Trans. Power Syst.*, **11**: 1996.
21. X. Guan *et al.*, An Genetic Algorithm for Power System scheduling, *Proc. 2nd Chinese World Cong. Intell. Control Intell. Autom.*,Xian, China, 1997, pp. 1105–1110.
22. V. Quintana, A. Chikhani, A stochastic model for mid-term operation planning of hydro-thermal systems with random reservoir inflows, *IEEE Trans. Power Appar. Syst.*, **PAS-100**: 1119–1125, 1985.
23. C. Li, R. Yan, J. Zhou, Stochastic optimization of interconnected multireservoir power systems, *IEEE Trans. Power Syst.*, **5**: 1487–1494, 1990.

24. G. Contaxis, S. Kavatza, Hydrothermal scheduling of a multireservoir power system with stochastic inflows, *IEEE Trans. Power Syst.*, **5**: 766–772, 1990.
25. S. Takriti, J. Birge, E. Long, A stochastic model for the unit commitment problems, *IEEE Trans. Power Syst.*, **11**: 1497–1508, 1996.

XIAOHONG GUAN
PETER B. LUH
Xian Jiaotong University
ERNAN NI
University of Connecticut
RENHOU LI
Xian Jiaotong University