Intelligent Unified Control of Unit Commitment and Generation Allocation

Final Report
EPRI Grant RP8030-13

John R. Birge, Samer Takriti, Erik Long

Department of Industrial and Operations Engineering
The University of Michigan

Abstract

Effective power system operation requires well-planned commitments of generating units in a planning phase combined with responsive real-time dispatching of power in an operational control phase. Typically, the unit commitment plan is determined by an optimization model that considers a single forecast for demand, fixed security constraints, and known unit availabilities. The operational controller considers actual random demand, varying security reserves, and unforeseen unit losses.

This project provides procedures that combine these tasks by integrating the planning and operating phases in a unified control framework. The emphasis in the planning phase is on considering random effects, contingencies and the values attributed to potential ending states. The emphasis on the operating phase is on considering longer term effects, recognizing conditions requiring corrective actions, searching for similar situations and potential actions and learning from both practical experience and simulated events. This report describes the general framework of this process, provides new methods for finding solutions that include uncertain loads and unit availabilities, and provides results that substantially improve upon the standard approach.

Contents

1 Introduction

Effective power system operation requires well-planned commitments of generating units in a planning phase combined with responsive real-time dispatching of power in an operational control phase. Typically, the unit commitment plan is determined by an optimization model that considers a forecast for demand, fixed security constraints, and known unit availabilities. The operational controller considers actual random demand, varying security reserves, and unforeseen unit losses.

This project provides procedures that combine these tasks by integrating the planning and operating phases in a unified control framework. The emphasis in the planning phase is on considering random effects, contingencies and the values attributed to potential ending states. The emphasis on the operating phase is on considering longer term effects, recognizing conditions requiring corrective actions, searching for similar situations and potential actions and learning from both practical experience and simulated events.

This report describes the general framework of this process, provides new methods for finding solutions that include uncertain loads and unit availabilities, and provides results that substantially improve upon the standard approach. In Section 1, we give a brief background on the problem. Section 2 describes the main open issues that we address. Section 3 describes the framework for intelligent situation recognition and action selection. Section 4 describes how this search and identification relates to an overall optimization framework for integrative planning and real-time control, called *match-up scheduling*. Section 5 describes the solution procedure in detail. Section 6 gives background on formulating the basic unit of our approach, a real-time subproblem solution. Section 7 provides the full formulation, which includes several deterministic problem solutions described in Section 8. Section 9 discusses the input generation which employs a fuzzy neural network approach. Section 10 gives our results using data from the Michigan Electric Power Coordination Center. Section 11 presents a summary of the work.

2 Brief Description of Problem

Decision making in power systems, as well as other systems, can be described in terms of the length of period directly affected by the decision. In the long run (measured in years), aggregate decisions such as capacity in terms of the number and type of power generation units are considered. In the medium term (measured in days, weeks or months), decisions are made about how generally to employ the capacity that is available. In the short run (seconds to hours), the allocations of generating capacity are made to meet demand.

This project focused on the integration of the second two areas - the coordination of planning and control decisions and the direct consideration of situation specific circumstances and the possibility for future random events. We use the terms, *unit commitment* and *generation allocation, or economic dispatch*, to describe these two areas.

2.1 The Generation Allocation Problem

The generation allocation (or *economic dispatch*) problem is a real-time control problem which allocates a received demand load, d, among n operating units such that the total cost of satisfying the load is minimized. Let F_i and x^i be the cost rate and output respectively from unit i, then the allocation problem is given by

$$\begin{aligned} Minimize \ & \sum_{i=1}^{n} F_i\left(x^i\right) \\ s.t. \ & \sum_{i=1}^{n} x^i = d + L, \quad \mathcal{P}(1) \end{aligned}$$

where L, the transmission loss, is a function of network current and impedances, often approximated quadratically with parameters, B_{ij}, by

$$L = \sum_i \sum_j x^i B_{ij} x^j + \sum_i B_{i0} x^i + B_{00}. \quad (1)$$

The minimum cost operating condition is determined from the solution of *coordination equations*,

$$\frac{dF_i}{dx^i} + \lambda \frac{\partial L}{\partial x^i} = \lambda \quad \forall i \tag{2}$$

which can be rearranged as

$$\left(\frac{1}{1 - \frac{\partial L}{\partial x^i}} \right) \frac{dF_i}{dx^i} = \lambda \quad \forall i. \tag{3}$$

The term in the parenthesis, known as the *penalty factor* for bus i, is typically very close to unity. This fact is exploited in computing the minimum cost solution—the penalty factor is computed *prior to* performing economic dispatch and is assumed constant during the solution of coordination equations. Once generator powers are obtained, the penalty factor is recomputed and the cycle is repeated, if necessary. Security limits on power generation, such as $\underline{x}^i \leq x^i \leq \overline{x}^i$, can be easily incorporated.

The methodology can be extended to multi-area power pools by coordinating the dispatch among area members in such a way that the cost for the pool as a whole is minimized. Let $\mathcal{A}$ be the set of areas within the power pool. Suppose x^a units of power generation in area a will cost $F^a\left(x^a\right)$ in operating expenses. The excess power interchange between area a and other areas, E^a, represents a major coordination decision. The coordination problem is given by

$$\begin{aligned} Minimize \sum_{a \in \mathcal{A}} F^a\left(x^a\right) \\ s.t.\ \ x^a = d^a + L^a + E^a \quad \forall a \in \mathcal{A}. \quad \mathcal{P}(2) \end{aligned}$$

Power pool members have ties with the external system, in which case, the net pool interchange, $\sum_a E^a$, has to be allocated among various members. A major difficulty in solving the coordination problem $\mathcal{P}(2)$ is determination of loss terms L^a's, which are typically computed by a tie line model using inter-area matrix sensitivity coefficients (Kirchmayer [23]; Happ citehapp). Also, highly differing response characteristics of hydro, nuclear and fossil generation units makes it difficult to regulate the megawatt output and overall system frequency simultaneously. Chowdhury and Rahman [15] provide a discussion and extensive references dealing with economic dispatch and its relationship to power flow and automatic generation control(AGC).

The generation allocation problem gives the minimum cost allocation for a received load. As the load changes, the solution process is repeated and generator outputs are adjusted. This method of solving a *sequence of static problems* has several limitations. It does not recognize the inherent limitations of generation units, such as the finite response rates of boiler and combustion equipment. The solution is generated independent of current operating points as if power units can respond instantaneously to changes. The sequence of successive target points may be infeasible to reach and, in fact, may even violate security constraints based on first order changes. Besides these problems, the static approach will not give a minimum cost solution due to several reasons. The cost rates are based on steady state values for fuel consumption while in reality, generator units are always in transient state, trying to track the successive target solutions. Moreover, the formulation does not account for additional fuel consumption caused by the transient off-nominal values of pressure and temperature encountered during the process of passing from one operating point to another. Finally, the fast ramp up and slow down resulting from a sequence of static solutions give rise to higher emission and increased cost due to thermodynamic irreversibility.

The aim of our work has been to develop dynamic dispatch methods which can deal with a high rate of load variation. We take into account the current operating point to discourage deviations using a smoothing model. We also incorporate rate-based security constraints and available information about future load such as daily variations, short-term weather forecasts and cyclicity. Our models produce gradient information that incorporates the uncertainty of the future as well as the current state. We may also incorporate additional factors such as CO_2, SO_x and NO_x emission so that policies that minimize particulate and thermal emissions can be obtained. These additional constraints do not change our fundamental model and may be of particularly high concern in the future (see, e.g., Prince [33]).

2.2 The Unit Commitment Problem

The objective of unit commitment is to schedule or "commit" a utility's generating units over the near future (a day to a week) such that the predicted demand over this period is met at a minimum operating cost. In addition to meeting the expected load, the unit commitment problem must take into account the spinning reserve requirements, minimum uptime and downtime of generation units and other technical limitations such as limits on power output.

Consider a unit commitment problem for the next T periods, each of length Δ (typically, an hour, but open to state-specific modification as we discuss below). Let $d(\tau)$ be the predicted demand profile as a function of time. For now, we disregard all potential uncertainties regarding predicted demand. The average demand in period t, is given by

$$d_t = \frac{1}{\Delta} \int_{(t-1)\Delta}^{t\Delta} d(\tau) d\tau \tag{4}$$

To cope with extreme demand situations, we also define the maximum and the minimum demand in any period

$$\overline{d}_t = \underset{(t-1)\Delta \leq \tau \leq t\Delta}{Max} \{d(\tau)\}$$
$$\underline{d}_t = \underset{(t-1)\Delta \leq \tau \leq t\Delta}{Min} \{d(\tau)\}. \quad (5)$$

Suppose the utility has a set $\mathcal{U}$ of generating units; each unit $i \in \mathcal{U}$ has a different economic characteristics and technical limitations. Let x_t^i be the power output of unit i in period t and $F_i(x_t^i)$ the operating expenses for maintaining this level of output over duration Δ. Let L_t be the transmission loss and E_t the excess power interchange from other units in period t. The power balance equation can be written as

$$\sum_{i \in \mathcal{U}} x_t^i = d_t + L_t + E_t \quad t = 1, \ldots, T \tag{6}$$

Let indicator u_t^i represent the status of unit i in period t, defined as

$$u_t^i = \begin{cases} 1 & if\ unit\ i\ is\ up\ in\ period\ t \\ 0 & if\ unit\ i\ is\ down\ in\ period\ t. \end{cases}$$

Each generation unit has upper and lower limits on output,

$$u_t^i g^i \leq x_t^i \leq u_t^i G^i \quad t = 1, \ldots, T; \ \forall i \in \mathcal{U}; \tag{7}$$

The running units must satisfy the spinning reserve constraints and extreme demand:

$$\sum_{i \in \mathcal{U}} u_t^i G^i \geq \overline{d}_t + L_t + R_t \quad t = 1, \ldots, T \tag{8}$$

$$\sum_{i \in \mathcal{U}} u_t^i g^i \leq \underline{d}_t. \quad t = 1, \ldots, T \tag{9}$$

The minimum uptime and downtime constraints must be added to avoid too many consecutive modifications in the status of a unit. Frequent startup and shutdowns are undesirable from a social as well as a technical point of view. Not only are they stressful to operators but thermal units also undergo a heating and cooling cycle coupled with pressurization and decompression of boilers, turbine chambers, etc. leading to reduced effective life of the generation units. Moreover, the emission of pollutants like, CO_2, SO_x and NO_x is particularly high during the transient period of startup and shutdown. To discourage frequent startups and shutdowns, we insist that a unit must stay *on* for at least T_i^{on} hours whenever it is started. Similarly, the unit must remain *off* for at least T_i^{off} hours whenever it is shutdown. To enforce such restrictions, let τ_{it}^{on} be the amount of time unit i has been on *continuously* up to (and including) period t. Similarly, let τ_{it}^{off} be the amount of time unit i has been off *continuously* up to (and including) period t. These variables are defined recursively as

$$\tau_{it}^{on} = u_t^i \left(\Delta + \tau_{i,t-1}^{on}\right) \tag{10}$$

$$\tau_{it}^{off} = \left(1 - u_t^i\right) \left(\Delta + \tau_{i,t-1}^{off}\right) \tag{11}$$

The minimum uptime and downtime requirements can now be given as

$$\left(\tau_{i,t-1}^{on} - T_i^{on}\right) \left(u_{t-1}^i - u_t^i\right) \geq 0 \tag{12}$$

$$\left(\tau_{i,t-1}^{off} - T_i^{off}\right) \left(u_t^i - u_{t-1}^i\right) \geq 0 \tag{13}$$

Finally, every time a unit is shut down, bringing it back into operation incurs an extra cost due to fuel waste, extra maintenance, and additional feed water and energy needed for heating. Since cooling of a unit is an exponential process, we assume that the cost of starting up a unit increases exponentially in standstill time $\tau_{i,t}^{off}$ with a characteristic time constant $\hat{\tau}_i$. Specifically, let s_i^0 be the fixed cost of shutdown and startup and s_i^∞ be the cost of cold-starting unit i then startup cost of unit i after a standstill time of $\tau_{i,t-1}^{off}$ is given by,

$$S_i\left(\tau_{i,t-1}^{off}\right) = Max\left\{0, \left(u_t^i - u_{t-1}^i\right)\left(s_i^0 + \left(s_i^\infty - s_i^0\right)\left(1 - e^{-\tau_{i,t-1}^{off}/\hat{\tau}_i}\right)\right)\right\} \tag{14}$$

The objective of unit commitment is to minimize the operating and startup costs over a horizon of T periods,

$$\underset{x_t^i, u_t^i}{Minimize} \sum_{t=1}^{T} \sum_{i \in \mathcal{U}} u_t^i F_i\left(x_t^i\right) + S_i\left(\tau_{i,t-1}^{off}\right)$$

$$s.t. \;\; (6) - (13) \quad \mathcal{P}(3)$$

Formulation $\mathcal{P}(3)$ is a 0–1 integer nonlinear program which may be quite difficult to solve. There is an extensive literature on the use of dynamic programming, integer programming, Lagrangian relaxation and various other relaxation/decomposition techniques applied to the unit commitment problem. We indicate only a few representative recent work here. Cohen and Sherkat [16] provide a more extensive review.

Bertsekas et al. [6] replace the integer variables corresponding to the status of each unit by their continuous counterparts.. The resulting convexified problem is then solved to yield a lower bound on the optimal cost of the original problem. From this they derive a feasible solution using heuristic rules. This is the basic model that we follow in our solution approach. We describe it in more detail in Section 5 below.

While the problem above is generally solved as a deterministic model with longer range cost, environment and reliability considerations, the actual operation of the system requires responses to uncertain conditions. An overriding factor in this operation is the system's security. Internal uncertainties in the generation units as well as external uncertainties such as lightning strikes, objects hitting transmission towers or human errors in setting controls can lead to unpredictable circumstances. The ability to avoid and recover from generation unit failures and transmission line outages is of prime concern for utilities. This concern leads to a priority in real-time control to respond effectively to varying conditions in any reasonable manner that matches loads and maintains security. Longer term financial and environmental objectives often become secondary at this level.

Secure and reliable operation of a power system involves close monitoring of line voltages, currents, and power flows, the status of circuit breakers and tap positions in the transmission line, and the frequencies and output levels of generating units. These data are continuously updated and telemetered to control rooms for contingency analysis and possible corrective measures. The heart of a security system is supervisory control and data acquisition system which allows operators to control circuit breakers and disconnect switches and transformer taps in case of overload or out-of-voltage situations. Typically, telemetered information are imperfect and require state estimation before potential action. In addition, many catastrophic events like generator tripping or cascading failures happen too fast for an operator to react. Defensive operation against such failures requires contingency analysis and fast action based on actual system data, state estimation procedures, and contingency event selection rules.

The result is that operators on the real-time dispatching level react in a different time frame to a different set of priorities from that of the engineers determining unit commitment from the planning problem in $\mathcal{P}(3)$. Their function requires integration of vast amounts of data about the situation facing the utility at that instant. Their decisions must instantaneously determine a reaction to changing conditions. The result is that much of the overall system performance depends on the operator's ability to integrate the planning objectives into the real-time reliable dispatching level. The planning system must also integrate these real-time considerations into its function in order to determine the higher level plans. The integration of these functions through intelligent control, situation identification, and through robust optimization procedures is the subject of this proposal.

3 Major Issues in Considering Uncertainty

The major issues addressed by our research are the impact of uncertainty on the unit commitment and generation allocation decisions and the value of the solutions obtained through those procedures. Our contribution is to provide a framework that allows the incorporation of real-time uncertainty into the operation of the full system.

a. *Unit Commitment Issues*

   1. **Uncertainty** The major difficulty in the unit commitment problem is how to respond to uncertainty in the load forecasts, startup/shutdown costs, emissions, and unit availabilities;
   2. **Quality/Evaluation** Another concern is how to evaluate plans generated by a unit commitment optimizer - how are costs, risks, and environmental concerns combined in making a decision and how can schedules be analyzed for other measures of quality.

b. *Generation Allocation Issues*

   1. **Uncertainty** In this real-time control framework, load variations and unplanned unit shutdowns must be dealt with - the questions concern how to construct responses that achieve goals given real-time dynamics and various triggering events;
   2. **Quality/Evaluation** The central issue in generation allocation is how well real-time decisions fit longer term goals in terms of costs and emissions. How can these factors be brought into the real-time framework and provide for quick responsive, yet economical controls?

c. *Coordination* - The difficulty in merging unit commitment and economic dispatch decisions is that the focus in each area is different. Unit commitment decisions are generally focused on longer term economic and environmental goals while real-time generation allocation decisions concentrate on dealing with unforeseen events and maintaining load. Our work is aimed at bringing more of the real-time aspects and potential costs into the unit commitment problem while also bringing more long term economic and environmental concerns into the the generation allocation problem.

To deal with the issues outlined above, we use the relevant information from the situation as presented to give operators the information needed to make decisions that are reliable but that are also consistent with higher level goals of efficient and environmentally sound operation. We will determine what situations require response not forecast in the unit commitment plan, how to categorize those situations and measure their similarities, how to evaluate actions made for varying situations, how to learn from actual events and from simulations, and how to present operators with ranges of choices, values of those choices, for actions in varying situations.

In the next section (3), we describe the basic intelligent control methodology. The following section (4) goes into more detail of the problem and shows more specifically the role of artificial intelligence and optimization in this control and planning problem. Section 5 describes the actual implementation.

4 Situation Recognition and Action Selection through Artificial Intelligence

The basic areas for intelligent control involve the following elements.

- *Intelligent Monitoring* - for load and generation units
- *Diagnosis* - based on past history and stored knowledge base of reference load profiles and actions
- *Information Display/Interface* - in a most suitable manner for human action; abstraction and referencing
- *Training* - of new operators together with a simulator

For monitoring and diagnosis, the major motivation for artificial intelligence is to characterize situations and to develop a metric for measuring similarity among situations. We wish to identify situations as beyond the

In general, adaptive control systems provide a framework for the control of complex systems (see Figure 1). The main components of such systems are: the *signal identification* stage which analyses the input signal to determine its properties; the *identifier* stage which identifies the state of the process by examining the instantaneous input and output characteristics of the process and by initiating specific test signals; the *decision computer* which receives the information from both the identifier stage and the signal identification stage and determines the required controller characteristics; and the *implementation* and *controller* module which interprets the decision and actually implement it.

In the power system, the signal corresponds to inputs from across the network on loads, power production, line voltages, and frequencies. The identification may also identify characteristics of change from the signal, weather input, and other conditions affecting the system state. The identifier considers both the input and output through monitoring of power outputs, responses and ramping rates. The decision computer is currently the operator but will consist in our model of operator and computer suggested options. The implementation and controller functions will include automated responses as well as responses that may include operator resposne. In general, we will allow for operator intervention and will concentrate on developing the decision computer to provide relevant information, informed sets of options, and evaluations of system state and the given options.

The main need for incorporating such an adaptive scheme arises from the complexity of the regulated process. For simple processes for which analytic mathematical models exist, there is no need to incorporate a complex scheme – a simple negative feedback loop suffices. As the nature of the process becomes more complex, and possibly includes some nonlinear characteristics, the simple approaches that analyse the input data based on very simple criteria are not sufficient any more, since the process exhibits a behavior of different characteristics at different ranges of operation (situations). In this case, the state of the process itself has to be identified (top-down) in order to determine the controller characteristics. The controller no longer has constant features but is expected to change with the process state. This capability increases the range of possible applications of adaptive control systems.

The success of adaptive control systems, is nevertheless, limited, and it traditionally deals with applications in which a small set of well defined signals control a process. All the information is in measurable signal form, and there is no heuristic or symbolic information incorporated into the identification or decision process. For an effective power system operation which needs to intelligently react to signals from the environment, a multitude of complexities is added to the problem, in particular No longer can the processes and tasks be defined in input/output the size of the raw input data becomes too large for control systems to handle in the traditional way. Another problem is quantification of information, and definition of normalized difference measures for different parameter values so that the 'distance' from a desired state can be measured, and used in the decision.

Despite difficulties in *analysis* of incoming signals and determining the optimal (or correct) reaction, simulations that provide a reliable *synthesis* of the result of a particular situation and reactions. With these systems, the incoming signals expected in a particular situation could be synthesized. The ability to synthesize 'input' information introduces new possibilities for intelligent control: the real incoming data is compared to synthesized information based on a priori knowledge, and the deviations between the two establish the information the system analyzes. This mode is referred to as the *verification* mode of data interpretation. This mode of operation is in contrast to the discovery mode in which information is sought from the sensory data directly, regardless of the state of the system. In this approach, a system simulation is used in the regular operation of the system. The deviations between anticipated signals and actual signals guide the reaction.

The main lesson to be learned from adaptive control is that as the processes become more complex, one cannot rely exclusively on the input data to determine an immediate response. The goal of the system is not simply to analyze its input data, but rather to identify its state. This task should be performed by assimilating new information (from sensors, or from remote stations) with the current state of the system (available knowledge). This approach brings up the problem of the amount of information in the combined state of the system, and requires the identification of the relevant portions of the state at any given time.

This problem is the Achilles heel for many complex computer systems which are expected to maintain more and more knowledge. Although the speed of computer systems is still constantly increasing, the amount of the necessary information we want these systems to consume productively is growing at an even higher rate. The problem this knowledge-explosion creates is how to access the proper information when it is needed. The problem could be stated in the following terms: how do we bring the necessary information 'closer' to the system so that it finds it with minimal delay when it needs it? For example, an overload of a particular generator unit may require specific knowledge which may contain previous similar experiences, simulation results, or simulation procedures. This information could be totally irrelevant to the system's other states of operation. The system, therefore, need not be bogged by this knowledge when it is irrelevant, yet, when it is relevant, it needs to have fast access to it. In our detailed description below, we refer to this task as isolating the disruption and determining a match up point.

We claim that power system operation has a set of known tasks to perform and a set of scenarios in which the tasks are performed. These restrict the knowledge expected to be needed at any given time. It is this expectation that allows us to resolve this knowledge explosion problem. We use the Knowledge Caching methodology for knowledge management. This methodology allows fast retrieval of information from large knowledge bases by exploiting locality of information. Knowledge Caching supports a model hierarchy of local and global models, it semantically clusters information, it defines relations between knowledge clusters, and it uses relations and current system's state to shuffle information between model hierarchies.

Much of the work in Knowledge Caching has focused on vision systems (see Roth and Jain [35, 36]). The state of the power system can also be viewed as large complex image with each input corresponding to a pixel in an image. In this way, vision based caching applies to more general situations. A key to this type of system is to fit the queries and analyses into the specific context of the power system domain.

Our focus in this effort has been to use the complete system history as a full knowledge base but to extract caches that correspond to similar situations as the current time. We then employ a fuzzy neural network model to derive from the active knowledge cache estimates on the future system evolution. These estimates then enter our optimization model constructs. In the next section, we describe the areas where these input signals are used in our control methodology.

5 Robust Planning Decisions and Match-Up Scheduling

Our control process can be described in a general framework called *match-up scheduling* that has previously been applied in manufacturing contexts (see Bean et al. [4]), but without the intelligent indentification and decision mechanisms we propose here. In match-up scheduling, a cyclic pre-schedule, based upon maximum plant uptime over downtime characteristics and startup costs, is created off-line. In case of disruptions, the operator takes a sequence of match up actions such that the operation returns to the normal pre-schedule at some future time. The choice of match up action such as unit shut down or turn up, emergency power interchange from power pool, etc., depends upon the severity of disruption. The economic dispatch and unit commitment problems are resolved for each candidate matchup times and a minimum cost solution is chosen.

This procedures continues in real-time to determine generation allocation considering changing weather and load forecasts, unscheduled unit outages, and unforeseen power transactions. This on-line real-time system will be a knowledge-based intelligent controller which can relieve the operator of repetitive responses. The operator will act at the highest level of control overseeing the control system and responding to unique situations by overriding the proposed controller.

The goal behind match-up scheduling is to combine the advantages of a static optimization procedure to produce schedules that consider future events with the benefits of a dynamic procedure that can rapidly respond to condition changes. The resulting schedule should then address global concerns with flexibility and robustness under system disturbances. This project will explore the value of this procedure by developing analytical models that can assess match-up capability in general environments and by testing the method using simulations of practical scheduling problems. The analytical results will concentrate on bounds that can be used to measure match-up performance generally. The experimental results will be used to verify the analytical models and to provide evidence of match-up performance in practice.

Match-up scheduling begins with the definition of a pre-schedule. In our terms, this function is the unit commitment problem. The preschedule can be determined with several objectives including economic costs, risks, and environmental factors. It also considers demand and unit availability uncertainty explicitly to provide extension over the forecast model in $\mathcal{P}(3)$. The pre-schedule of unit commitment then includes options to follow with relatively small demand and supply variations. When the system state goes outside this window considered in unit commitment, we say that a disruption (for example, loss of large unit, large load, frequency or load rate change) has occurred. At a disruption, the match-up policy is to choose a point at which to match up with the pre-schedule. The match-up point includes both the units that must be potentially re-scheduled and the times on those units when match up will occur. This subproblem represents a (potential) change in generation allocation policy. The actual control must be applied in real time.

Turnpike theory provides some economic justification for the match-up technique. In Bean et al. , it is shown that under known conditions, the match-up solution can come as close as desired to an optimal solution by choosing a suitable match-up point horizon. For multiple disruptions that are sufficiently well-spaced apart, match-up scheduling can give bounds on the average cost difference between a match-up solution and an optimal solution. Birge and Dempster [11] also show that the match-up/turnpike results hold in a more general, stochastic context. The main practical value of this analysis is in an ability to bound the effects of following a given real-time dispatching strategy and to provide input on the value of searching for new plans.

The definition of a match-up point determines the subproblem on which to apply solution procedures. This is one of the six basic steps in match-up scheduling:

1. Determine a pre-schedule (unit commitment and normal mode generation allocation schemes);
2. Identify a disruption and classify it;
3. Define a match-up point and subproblem;
4. Define objectives for matching up;
5. Determine allocations and possible new unit commitments for subproblem;
6. Update the remainder of the unit schedule after subproblem solutions if warranted.

These steps essentially represent how the power system is actually controlled now. The unit commitment decisions are followed generally until some disruption occurs. Most generation allocation decisions follow a set pattern first to meet the current load and then to dispatch with a generally static priority rule corresponding to a solution to $\mathcal{P}(1)$. The match-up framework provides a link between these problems and some justification for limited unit re-commitments at the real-time level. The critical areas of uncertainty and overall schedule evaluation are, however, the focus of this project. We next describe how they enter into each match-up phase given above.

5.1 Unit Commitment - Pre-Schedule Selection

The formulation of unit commitment in $\mathcal{P}(3)$ ignores the central difficulty of uncertainty. It also does not include various aspects that we would like to address in terms of risks (attitude toward uncertainty) and environmental factors. Some models have considered uncertainty explicitly, but these have generally focused on high level, long term capacity planning problems (e.g., Bloom [13], Borison et al. [14], Dantzig and Glynn [17], Louveaux and Smeers [28]). In our near term decisions, however, we consider the set of generation units fixed and focus on commitment of specific units.

Considering uncertainty is especially important to utilities because of new environmental regulations. For example, due to annual limits on sulfur dioxide emissions and trading of pollution rights, an options market is being created to allow utilities to hedge against excessive emissions. These options would never enter the solution of any deterministic model because they would always be unnecessary (one would either sell or buy just enough to meet known emission levels). This inability of deterministic models to capture hedging behavior can lead to poor planning (see, for example, Birge [9]) and costly future penalties.

One example of solving a unit commitment problem under uncertainty, involving pump storage, is in Pereira and Pinto [31] who used a method similar to that given in Birge [7]. Our basic model follows these developments. To ease our development here, we assume that the power flow equations and other relationships can be linearized (as in Deák et al. [18] or Granville and Pereira [20]). In this case, the stochastic version of $\mathcal{P}(3)$ becomes a multistage stochastic mixed integer linear program. We describe this model in a general form where x_t represents all variables in time period t. The actual decision vector depends on unknown outcomes. We denote these outcomes as random elements $\omega \in \Omega$. In this way x_t is really a function of the outcome, $x_t(\omega)$. We write this in bold-face notation, $\mathbf{x}_t$. We also divide Ω into components that happen in each period so that $\Omega = \Omega_1 \times \cdots \times \Omega_T$. We suppose that $\Omega^t = \{\omega^t = (\omega_1, \ldots, \omega_t) | \omega_i \in \Omega_i, i = 1, \ldots, t\}$, which represents the past at time t. In this way, we can technically characterize all events up to time t by measurability with respect to the Borel field, $\mathcal{F}^t$, defined by Ω^t. This yields the general stochastic program to find $\mathbf{x} = (x_0, \mathbf{x}_1, \ldots, \mathbf{x}_T)^T$ to

$$\begin{array}{lllll} \min c_0^T x_0 & +E_\Omega[\mathbf{c}_1^T\mathbf{x}_1^T + & \cdots & +\mathbf{c}_T^T\mathbf{x}_T^T] & \\ \text{s. t. } A_0 x_0 & & & = b_0, & \\ & \mathbf{B}_{t-1}\mathbf{x}_{t-1} & +A_t\mathbf{x}_t & = \mathbf{h}_t, & \\ & & t = 1, \ldots, T; & a.s., & (\mathcal{P}(4)) \\ & & \mathbf{x}_t & \geq 0, & \\ & & t = 1, \ldots, T; & a.s., & \\ & \mathbf{x}_t & -E_{\Omega_t}[\mathbf{x}_t] & = 0, & \\ & & t = 1, \ldots, T; & a.s., & \end{array}$$

where we assume that all the constraints in $\mathcal{P}(3)$ are represented in the first three sets of constraints. The last constraint set, $\mathbf{x}_t - E_{\Omega_t}[\mathbf{x}_t] = 0$ almost surely, means that $x_t(\omega)$ must be the same for all ω that appear the same up until time t. In other words, our action at t cannot *anticipate* the future. For this reason, it is called a *nonanticipativity* constraint.

Several questions arise in the formulation of $\mathcal{P}(4)$. We describe them as:

- Choice of periods and horizon
- Choice of probability distribution and approximation
- Choice of objective
- Solution approach
- Solution evaluation

5.2 Period Selection

First, the model in $\mathcal{P}(4)$ assumes a given set of periods and period lengths. This decision may have a significant effect on the result since we assume unit commitment decisions are constant within each period. This is one of the subjects of our intelligent search over system states and the quality of prior solutions. The result is:

KC1 :time periods and period lengths to define a unit commitment model with best solution quality subject to overall size restrictions and given the current system state.

Some results concerning (KC1) are that overall bounds on aggregating dynamic decisions into fixed time periods are available (Birge [8]). These bounds play a role in the solution evaluation phase as well. In our analysis of MEPCC data, we found that hourly periods and a weekly time horizon provide effective results. For finer divisions we found that demand branches at peak and low load values were contributors to the hedging types of solutions that we developed. Finer subdivisions for scenarios did not improve the results significantly.

5.3 Distribution Selection

Given fixed time periods in $\mathcal{P}(4)$, the model in $\mathcal{P}(4)$ still presents several difficulties. For example, what is the distribution of ω and, if too complex for inclusion in an optimization problem, how can that distribution be adequately represented? The "distribution" corresponds to what one might expect in terms of loads and unit characteristics given current state information. Evaluating a complete distribution is probably not possible given the level of uncertainty in the problem. Instead the best one can probably do is to represent certain aspects of the problem using limited information (mean forecasts and variances) and other state information represented in the knowledge cache. Various methods for doing this are described in Birge and Wets [12].

A difficulty with these methods is that they generally require many alternative future outcomes or *scenarios*. In generally we would like to find a limited set of scenarios so that at each period t, we can partition Ω^t as $\Omega^t = S_{t,1} \cup \cdots \cup S_{t,\nu_t}$. Each $S_{t,i}$ represents possible outcomes up until time i. From each i, we might have several outcomes in time $t+1$ which we call *descendants* of i, written as $\mathcal{D}_{t+1}(i)$. In this way, $S_{t,i} = \cup_{j \in \mathcal{D}_{t+1}(i)} \{\omega^t | (\omega^t, \omega_{t+1}) \in S_{t+1,j}\}$ so that the partitions are consistent from one period to another. The first task in selection the distribution is then to define these scenarios.

Our method for scenario generation is built on our fuzzy neural network outputs combined with information on the reliability of the generation units. The fuzzy neural network model uses the historical data within the knowledge cache and current information to define a set of similar histories. These are weighted to produce a small set of candidate load scenarios. These scenarios are in turn modified by adding power interruptions such as a plant's tripping off-line.

The weights on each scenario are equivalent to assigning a probabilty, $p_{t,i} = P[S_{t,i}]$, to each outcome and some expectations, $E_{S_{t,i}}[\mathbf{c}_t] = c_t$ and $E_{S_{t,i}}[(\mathbf{h}_t, \mathbf{B}_t)] = (\bar{h}_{t,i}, \bar{B}_{t,i})$. If we construct the scenarios and assess these probabilities, the problem is then to find:

$$\begin{array}{llcc}
\min c_0^T x_0 & + \sum_{i=1}^{\nu_1} p_{1,i} c_1^T x_{1,i} + & \cdots & + \sum_{i=1}^{\nu_T} p_{T,i} c_T^T x_{T,i} \\
\text{s. t. } A_0 x_0 & & & = b_0, \\
 & \bar{B}_{t-1,i} x_{t-1,i} & + A_t x_{t,j} & = \bar{h}_{t,j}, \\
 & & t = 1, \ldots, T; & i = 1, \ldots, \nu_{t-1}; \\
 & & & j \in \mathcal{D}_{t+1}(i); \\
 & & x_{t,i} & \geq 0, \\
 & & i = 1, \ldots, \nu_t; & t = 1, \ldots, T.
\end{array} \qquad (\mathcal{P}(5))$$

One result from Birge and Wets is that $\mathcal{P}(5)$ actually gives a lower bound on the value in $\mathcal{P}(4)$ assuming that the probabilities and expectations can be determined from the available information. In general, this choice of scenarios and probabilities should try to match as much information about potential outcomes as possible. The results is a knowledge cache:

$KC2$:*a set of scenarios and assigned probabilities that match as closely as possible all information about possible future outcomes.*

5.4 Objective Selection

In the deterministic case, the objective generally reflects known costs and may include constraints on known emissions. Since none of these quantities is actually known, many uncertainties arise in the objective definition. The uncertainties also cause the problem to have some associated risk. In the linear case, our model actually assumes risk neutrality. This may not be a prudent decision since some high cost, low probability events may not be considered as much as might be desired.

We, therefore, need to consider whether other objective functions are worthwhile. This function is part of the intelligent modeler. The decision may not be to pick one specific objective but may include several and may also adjust the constraints, for example, on emissions. The result is that we can choose several objective-constraint combinations for modeling in $\mathcal{P}(4)$ (with perhaps nonlinearities) and can report more distributional information about costs than the expected utility framework used in $\mathcal{P}(4)$. We conduct this task by building a table of the possible results (see Tables 2 and 3). In this way, the operator sees the entire realm of possible outcomes under each selected policy. This results in the following cache.

$KC3$: *sets of objectives and constraints that model sets of preferences as closely as possible and pick out key aspects of constraint and objective distributions for reporting to decision maker.*

5.5 Solution Approach

The solution of $\mathcal{P}(5)$, even with linearization and finite numbers of future scenarios, may be quite difficult due to the integer variables corresponding to unit's being up or down and going on or off in each period. In some power system models without pump storage (see, e.g., Louveaux and Smeers [28]), this creates few problems because the periods can be viewed independently, leading to a property called *block separable recourse* (Louveaux [27]). This attribute is also exploited in Avriel et al. [2]. Block separable recourse enables the model to be solved with integer variables only in the first period so that efficient decompositon methods can be applied.

In the unit commitment problem, especially with pump storage, no separation of periods is possible because demand between periods is highly correlated (due, for example, to weather) and because pump storage level is carried over between periods. The result is that the value function of dynamic programming loses convexity properties and makes decomposition quite difficult (see, e.g., Stougie [39]). One approach is to solve the problems for pump storage and thermal generation in separate phases, as in Bertsekas et al. [6] who use lagrangian relaxation on the separable thermal problem, and in Bannister and Kaye [3] who assume a known production cycle among thermal units.

Our approach maintains simultaneous pump storage and thermal commitment decisions by developing solution *patterns* corresponding to feasible cycles of varying sets of units. Each pattern y^i, $i \in I$, corresponds to fixing some set of integer variable levels within each $x_{t,i}$. I This allows us to decompose the problem as in Birge [7], Louveaux [27], and Pereira and Pinto [31], and enables parallel computation. Techniques such as lagrangian relaxation also allow bounds on the solution values and determine when the solution process can stop. We give the details of our approach in the following section. In this procedure, we construct the cycle pattern that enters the knowledge cache:

$KC4$: *construct patterns of unit cycles under various scenarios that lead to efficient and precise models.*

5.6 Solution Evaluation

Given a solution of $\mathcal{P}(5)$, a first type of evaluation is how far the optimization result might be from a true optimum. We accomplish this through standard dualization procedures and other methods based on the problem structure. Other solution characteristics concern the model as given in $\mathcal{P}(5)$ and possible changes from it. For example, we mentioned earlier that $\mathcal{P}(5)$ as presented actually represents a lower bound on the value of the problem in $\mathcal{P}(4)$. Upper bounds (representing the worst expected utility given known distribution information) can also be constructed using the techniques in Birge and Wets [12] and Birge [10].

We made other evaluations using simulations of the scenarios and application of the decisions made in solving $\mathcal{P}(5)$. The simulations include responses to disruptions as given by the real-time dispatching procedures which react to unexpected shutdowns. Overall, we compare our approach to methods that use a single forecast value. The results appear in Tables 2 and 3. For a system, the results then compose:

$KC5$ *key solution characteristics under different scenarios to use in matching risk attitudes, preferences and areas of concern.*

5.7 Generation Allocation

Given unit commitment decisions, the next phase of the intelligent power system controller is to determine generation allocation schemes. In general, we assume that these allocation schemes are also generated and evaluated within the knowledge base. They would represent different responses to changing conditions, such as using least marginal cost first increases as power increases. This static approach may not represent an optimal response to load inputs because it ignores the dynamics of the situation and generally does not include long run costs. The key task of the knowledge caching system is then to

$KC6$ *construct generation allocation control decisions as functions of system state and in response to real-time inputs.*

Since our system includes real-time knowledge into the solution of $\mathcal{P}(5)$, we actually accomplish $KC6$ by ensuring that we can solve problems sufficiently quickly to be able to affect the real-time generation control decisions. Through our use of a KSR parallel processor, we were able to achieve such rapid response. We have essentially followed the procedures in Gröwe and Römisch [21] by assuming a limited open-loop form of generation control with penalties paid in real-time for deviations from actual load. We close the loop by solving disaggregated (indeed continuous time) versions of $\mathcal{P}(4)$. Additional constraints appear in these dynamic models, however, because of limitations on ramping rates (e.g., $x_t^i - \rho_i \leq x_{t+1}^i \leq x_t^i + \tau_i$, where ρ_i is limit on power reduction in a short time interval and τ_i is limit on power increase). While fully optimal solutions may not be obtainable, we use knowledge gained from previous solutions to accelerate searches for decisions and to obtain robustness and security while also maintaining longer term goals. In this way, the knowledge system identifies

KC7 responses to changing conditions given unit commitments that respond as well as possible according to the criteria and priorities observed within the system.

We note that many factors may enter the system including time histories, rates of changes (even second order effects), weather conditions (and rate of change), and external connections.

The form of intelligent controls may at first be to use the system to determine longer range marginal cost inputs (considering unit commitments already determined and given emissions and including multiplier or dual variable information from the unit commitment solution) and to modify these costs according to system state. This would allow online evaluation of transactions to determine relative merits of sales or purchases in addition to determining the operation of units within the system.

Other more complex control responses (involving simultaneous changes in several units in response to varying inputs) may also be developed to reflect more accurately the linkages among all units. These types of responses would again be found through the knowledge base and search described in KC7. These responses would all assume no significant changes in unit commitment from that which was planned as part of some scenario in the unit commitment solution. However, this may not always be possible and leads to the next part of disruption and match up identification.

b. *Identify a disruption and classify it*

The system state is constantly updated and checked against the states considered in the unit commitment scenarios. The neural network continuously updates the forecast and compares with what has occurred. When the forecast scenario weights differ significantly from the original forecast then the match-up process begins with a new unit commitment solution. Other events, such as shutdowns, major line voltages, and frequency changes, also trigger this process. The resulting knowledge base:

$KC8$ *determines whether a disruption has occurred which was not planned in the unit commitment solution and the type of disruption for match up selection.*

This procedure will also involve choices of units according to which scenario used in unit commitment is most closely followed. The process therefore involves continuous monitoring and search for the closest scenario, selection of general control responses for that scenario, and evaluation of distance from the scenario.

Having identified a disruption, the next step is to decide when and how to match up with the original unit commitment.

c. *Define a match-up point and subproblem*

d. *Define objectives for matching up*

A key decision in match-up scheduling is the choice of the horizon over which to plan the match-up schedule. Since the match-up point depends both on the units and time, the horizon can vary according to the unit (short time for some and perhaps longer for others). One choice for determining a match-up horizon is to consider the uncertainty inherent in the schedule and to attempt to match up to pre-schedule times with some confidence. To obtain indications of subproblems in which disruption effects can be isolated, we find conditions under which solving the subproblem produces the same performance characteristics as solving the full problem. Our analysis with MEPCC data has shown that one or two-day windows often produce very similar results to the weekly schedule when the new solutions are constrained to match up with the previous solution's water storage after two days. This allows faster solutions when considering a short-term decision. The overall knowledge base then:

$KC9$ defines a match up point, objective criteria, and subproblem for solution after a disruption.

Having defined the match-up problem, the next step is to:

e. *determine allocations and possible new unit commitments for subproblem.*

The match-up scheduling method operates in real time in response to changing conditions. Given this situation, the subproblem solution procedure must provide good solutions quickly. Time consuming efforts to obtain optimal solutions are not indicated. We also measure the robustness of the procedures by varying implementation conditions and analyzing the *distribution* of the performance criteria and not just the *expectations*. Our results in Tables 2 and 3, for example, give the entire distribution of results over the selected scenarios.

Our general approach solves a disaggregated version of unit commitment with parameters determined in the problem definition phase. Since solutions must be obtained quickly, however, the general approach chooses a variety of solutions quickly. We attempted a variety of procedures for finding solutions quickly. The results indicate that a method based on *simulated annealing* (see, e.g., Kirkpatrick et al. [24]) appears to offer the most robust solutions. This procedure avoids the trap of many local improvement schemes by tending probabilistically toward a global optimum (under certain conditions). This is then our preferred method to:

$KC10$ search for quick efficient response patterns given disruption, state and match up problem definition, identify sets of patterns and choices for branching among pattern solutions for quick computation.

The remaining step in matchup scheduling is to

f. *update the remainder of the unit schedule after subproblem solutions if warranted.*

The solution of the match up problem may lead to changes beyond the match up horizon because of system constraints (minimum up and down time). We allow these changes to occur but again with a transfer of information to unit commitment and information from the unit commitment solution about marginal values of resources. At any time, full solution evaluation is always possible possible.

The value of this system is in its ability to work in real-time respond to practical situations and provide a means for hedging against uncertain events and for including long term effects into real-time decisions. To do this, we have used actual data from the MPECC. This gives us characteristics of the power system throughout Michigan and its ties to neighboring systems. In the following section, we describe our solution procedure and computational results for solving the unit commitment subproblems in more detail.

6 The Stochastic Unit Commitment Subproblem Solution

This project has investigated each of the areas above and developed a comprehensive model using the Michigan Electric Power Coordination Center data for testing. The emphasis has been on developing intelligent methods that can apply in real-time to the unit commitment and generation allocation problems. To make this description general, we assume n generating units and that the duration of the study horizon is T periods. A 24-hour horizon may be sufficient, but a longer horizon, one week or more, is often needed if pump or other storage units are considered. The state of each unit, i, at a time period, t, is represented by the 0-1 variable, u^i_t. A unit is on at time t if $u^i_t = 1$, and off if $u^i_t = 0$. The power output level at which unit i operates during a period, t, is $x^i_t \geq 0$. As above, the minimum and maximum operating levels for each unit, i, are g_i and G_i, respectively.

set equation number start

The cost function, F_i, of operating a unit, i, at a level, x^i_t, is assumed to be a convex quadratic function of x^i_t. This function measures the total fuel and maintenance cost associated with each output level, x^i_t, in the feasible operating range [37]. A start-up cost, S_i, is incurred whenever the state of a unit changes from zero to one. Also, a shut-down cost is incurred when the state of a unit changes from one to zero. In some systems, the start-up and shut-down costs are functions of how long a unit has been off or on [26]. We modify the cost function, F_i, of each unit, i, so that it takes into consideration the start-up and shutdown costs. We refer to this modified function by $f_i(x^i_t, u^i_{t-1}, u^i_t)$. Note that the function f_i is not continuous. Hence, it is non-differentiable and non-convex. When a unit is switched on, there is a minimum on-time requirement. As noted above, a similar constraint applies for the case when a unit is switched off.

The mathematical formulation of $\mathcal{P}(3)$ with these simplifications is:

$$\begin{array}{llcl} \min_{x,u} & \sum_{i=1}^{n} \sum_{t=1}^{T} f_i(x_t^i, u_{t-1}^i, u_t^i) & & \\ \text{s.t.} & \sum_{i=1}^{n} x_t^i & \geq & \tilde{d}_t, \end{array} \tag{15}$$

where $\tilde{d}_t \geq 0$ represents the total demand for electricity during period t. In addition to the previous constraints, each unit must satisfy the minimum on-time, minimum off-time, and minimum and maximum operating level constraints. The problem in (15) is a large-scale mixed-integer quadratic program.

Many approaches have been proposed to solve this problem [1]. They can be classified into branch-and-bound methods, dynamic programming, priority ordering, and the Lagrangian relaxation method. The first two techniques are satisfactory from the theoretical point of view, but they are practically intractable due to the large storage size required to implement them on a computer. The third approach is a greedy strategy and does not guarantee an optimal solution in general. The Lagrangian relaxation technique seems to be the most efficient, because it attempts to solve the problem indirectly by solving the dual problem.

Muckstadt and Koenig [30] appear to have first addressed the unit commitment problem with mathematical programming and to suggest a sound technique for solving it. They use a Lagrangian relaxation technique which decomposes the given problem into smaller subproblems. Each sub-problem corresponds to minimizing the cost of operating a generator in the electrical system over the study horizon. Dealing with individual generators simplifies the task of representing the constraints that depend on the state of the generator from period to period, such as the minimum on-time and minimum off-time requirements. Dynamic programming is then used to solve the subproblems, and a lower bound on the optimal cost of the primal problem is obtained.

Given this lower bound, Muckstadt and Koenig use branch-and-bound to enumerate all possible states of the system efficiently. At each node of the branch-and-bound tree, the states of some generators at certain time periods are fixed and the Lagrangian dual of the problem at that node provides a lower bound that helps in pruning the search tree. However, to maintain high efficiency, they do not allow more than a small number of Lagrange multiplier updates at each node. The update is performed using a subgradient method that approximates the steepest-ascent direction since the dual gradient is not unique at some points. The dual solution at some nodes may provide a feasible solution which further reduces the size of the search tree. This strategy—that is, Lagrangian relaxation coupled with branch-and-bound—is common in integer programming and is guaranteed to find an optimal solution.

The previous technique may fail in practice, as shown by the authors, due to the large number of nodes that need to be studied. Bertsekas, Lauer, Sandell, and Posbergh [6] use a similar Lagrangian relaxation technique, but, rather than using branch-and-bound, they update the Lagrange multipliers and resolve the problem. The process is repeated until the duality gap is small enough to reach an acceptable answer. To accelerate calculations, the cost function of each generator's sub-problem is approximated by a differentiable function. The approximate dual problem is solved using a quadratically convergent constrained version of Newton's method, which makes use of the gradient and the Hessian matrix of the approximate dual function.

Bertsekas et al.'s main contribution is the provision of an upper bound on the size of the duality gap when the number of generators is greater than the study horizon. Their bound is given by $\frac{2T(S^*+TC^*)}{\pounds^*}$ where T is the length of the planning horizon, S^* is the maximum start-up cost over all generators, C^* is the maximum cost of operating a generating unit for one period, and $\pounds^*$ is the optimal dual functional value. The previous result is consistent with a well-known property of Lagrangian relaxation: "The duality gap is typically quite small (in relative terms) if the number of separable terms is large, and in fact becomes smaller as the number of separable parts increases" [5].

A similar approach is suggested by Merlin and Sandrin [29]. They use the same Lagrangian relaxation technique to separate the problem. Pump-storage hydro units are treated as any other generator is treated. They then use the subgradient algorithm to update the Lagrange multipliers as proposed by Poljak [32]. Their numerical results are impressive given the computing power at that time. They solved a system with 172 units over a 48-hour horizon to within 0.42% in two minutes. These results may, however, be a consequence of using linear cost functions for the generators while other references use piecewise-linear cost functions.

For a further improvement, Zhuang and Galiana [40] provide a heuristic that can be used at termination of the dual maximization procedure if the resulting primal solution is infeasible. The main idea is to increase, with a moderate step size, the penalties on the violated constraints iteratively until a feasible solution is found.

In all of the above-mentioned studies, the focus was on solving the unit commitment problem given that the demand at any time period is known in advance. None of these papers addressed the stochasticity of the load on the system and how to treat the demand uncertainty in a satisfactory fashion. Clearly, the demand for electricity in (15), $\tilde{d}_t$, is a random variable that fluctuates according to the weather, the day of the week, the time of day, and many other factors.

In practice, the demand during each period, which is usually one hour, is estimated using the weather forecast and the data that already have been collected from similar periods in previous years. This estimation is made for each period in the problem horizon, which is generally one week. The approximated demands are then used to find an optimal strategy, which determines the status of each generator during each time period.

Since the actual demand at any point in time may be different from the approximated demand, experienced operators observe the changes in the demand and make the necessary real-time decisions. Typically, generators with the lowest cost per electricity unit are placed on an Automatic Generation Control to meet load and maintain frequency. In this case, the operator is trying to meet the demand at a low cost using a greedy strategy. After using the flat parts of the cost functions for the running generators, it may, however, actually be more beneficial to start another generator than to use the generator with the smallest slope. In some cases, all running generators may reach their upper limits, and still another generator may need to be started.

A widely used approach to handle demand uncertainty is to develop forecasts of both the average, d_t, and the peak demand, $\overline{d}_t$, for each time period. The unit commitment problem formulation is then altered so that it takes into consideration the worst case scenario, that is, the peak demand at each period. The result is:

$$\begin{array}{ll} \min_{x,u} & \sum_{i=1}^{n} \sum_{t=1}^{T} f_i(x_t^i, u_{t-1}^i, u_t^i) \\ \text{s.t.} & \sum_{i=1}^{n} x_t^i \geq d_t, \\ & \sum_{i=1}^{n} G_i u_t^i \geq \overline{d}_t. \end{array} \tag{16}$$

Planners require excess reserve capacity not only to guarantee the existence of enough generating capacity, but also to protect the system against the inability to satisfy the demand when generating equipment failures occur. Excess system capacity is often called spinning reserve [30] which is required to be at least 3% of peak load [19].

As noted above, the problem with this formulation rests in the mediation of loads. Our techniques for handling this appears below in Section 7. Improvements over the current strategies used in solving the classical unit commitment model are presented in Section 8. Section 9 describes how to model some of the factors that cause demand uncertainty using our formulation. We present some numerical results in Section 10.

7 Stochastic Unit Commitment Formulation

Rather than solving the unit commitment problem for the expected and the worst-case demand vectors, one can consider a set of possible vectors. Each of these demand vectors—or scenarios—represents a limited amount of information about the demand uncertainty. Each scenario is assigned a weight, P_s, that reflects the possibility of its occurrence in the future. In other words, the uncertainty about future demand is modeled by a number of deterministic subproblems, an approach called scenario analysis [34]. One hopes to discover similarities and trends by studying different instances of the subproblems and, eventually, to develop a "well-hedged" solution to the original problem.

The policy we are looking for must satisfy the constraint that if two different scenarios s and s' are indistinguishable at time period t on the basis of information available about them at time t, then the decision made for scenario s must be the same as that of scenario s'. The previous constraint is modeled by partitioning the scenario set at each time period into disjoint subsets that are called scenario bundles. Clearly, a bundle at time t is refined in subsequent time periods into smaller disjoint bundles [34].

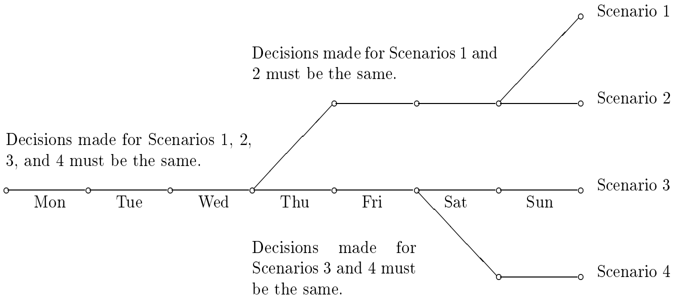

Figure 2: Scenario Bundles

To clarify the previous concept, assume that the unit commitment problem was solved for S demand vectors, d^s, $s = 1, \ldots, S$, which resulted in a three-dimensional array representing the status of each unit at each time period under each scenario, $u_t^{i,s}$. The system administrator needs to make a decision concerning the status of each unit, i, during the first time period based on the solutions obtained. In general, the values of $u_1^{i,s}$, $s = 1, \ldots, S$, are not equal. An optimal decision cannot be made without reformulating the problem sothat the decisions $u_1^{i,s}$ are the same for all $s = 1, \ldots, S$ and $i = 1, \ldots, n$.

One must not only consider the first time period but must also take into account all subsequent bundles that could affect the decisions made throughout the study horizon. Two scenarios are members of the same bundle, Ω_k, at time period t if both of them have the same demand values for all time periods $1, \ldots, t$. Note that each scenario is a member of exactly one bundle at each time period, which motivates the notation $B(s,t)$. If two scenarios, s_1 and s_2, are members of the same bundle at time t, then their bundles in time periods $1, \ldots, t-1$ are also the same. In other words,

$$B(s_1,t) = B(s_2,t) \Rightarrow B(s_1,\tau) = B(s_2,\tau),\ \tau = 1,\ \ldots,\ t-1.$$

Mathematically, a bundle at time period t is represented as a constraint on the decision variables, $u_t^{i,s}$, of its scenarios. Adding the bundle constraints results in a large-scale mixed-integer quadratic program that combines S unit commitment problems together. The objective function is to minimize the weighted sum of the objective functions of the smaller problems; that is, to minimize the expected cost over all possible scenarios. Here is the mathematical formulation:

$$\begin{array}{ll} \min_{x,u} & \sum_{s=1}^{S} P_s \sum_{i=1}^{n} \sum_{t=1}^{T} f_i(x_t^{i,s}, u_{t-1}^{i,s}, u_t^{i,s}) \\ \text{s.t.} & \sum_{i=1}^{n} x_t^{i,s} \geq d_t^s, \end{array} \tag{17}$$

and

$$B(s_1,t) = B(s_2,t) = \Omega_k \Rightarrow u_t^{i,s_1} = u_t^{i,s_2} = c_k^i,\ i = 1, \ldots,\ n.$$

Note that our mathematical model of (17) is a generalization of the traditional unit commitment formulation of (16). For instance, our scenario set can include the most-likely scenario, d_t, the worst-case scenario, $\overline{d}_t$, in addition to other scenarios that may be needed to model the stochasticity in the system. Since the bundle constraints are enforced at each time period, the demand and reserve constraints are satisfied throughout the planning horizon.

We solve program (17) in a Lagrangian relaxation type of technique. A multiplier, $\mu_t^{i,s}$, is associated with the bundle constraint on each variable $u_t^{i,s}$. The corresponding penalty term, $\mu_t^{i,s}(u_t^{i,s} - c_k^i)$, is added to the objective function. The target value c_k^i is the same for all scenarios that share the same bundle and is assumed to be the weighted average of the decisions made:

$$c_k^i = \frac{\sum_{s:B(s,t)=\Omega_k} P_s u_t^{i,s}}{\sum_{s:B(s,t)=\Omega_k} P_s}.$$

Starting from a set of penalties, S unit commitment problems are solved. If the resulting solutions satisfy the bundle constraints, then we stop with an admissible implementable policy or a feasible strategy. Otherwise, the penalties are updated and the process is repeated. We call the previous process progressive hedging [34]. Since the program in (17) is not convex, there is no guarantee that the sequence generated by the previous algorithm is convergent. Also, in case of convergence, the feasible policy obtained at termination may be a local minimum for the problem. However, numerical experience shows that the previous method converges rapidly and that the solution obtained is satisfactory. To get better policies, different starting penalties and different updating strategies should be used.

Note that the dual problem resulting from relaxing the bundle constraints is separable into S disjoint problems, each of which is a deterministic unit commitment problem of the form[1]:

$$\begin{array}{ll} \min_{x,u} & \sum_{i=1}^{n} \sum_{t=1}^{T} f_i(x_t^{i,s}, u_{t-1}^{i,s}, u_t^{i,s}) \\ \text{s.t.} & \qquad\qquad \sum_{i=1}^{n} x_t^{i,s} \;\; \geq \;\; d_t^s . \end{array} \tag{18}$$

The previous remark encourages a parallel implementation of this technique. Each deterministic problem, as in (18), is solved individually on a designated processor. If the solution obtained is infeasible, the penalties, $\mu_t^{i,s}$, are updated and each of the resulting new problems is again sent to a processor. Since most of the computational effort is used to solve the deterministic unit commitment problem (18), we describe various techniques to speed up the calculations in the following section.

8 Deterministic Unit Commitment Problem

To make the program in (18) separable, a Lagrange multiplier, $\lambda_t \geq 0$, is associated with each of the constraints $\sum_{i=1}^{n} x_t^i \geq d_t$. This choice of relaxation decomposes the problem into n single-generator subproblems. Constraints that depend on the change in the state of a generator from period to period, such as minimum up-time and minimum down-time, become easy to implement.

The Lagrangian dual problem has the form

$$\max_{\lambda \geq 0} \mathcal{L}(\lambda)$$

where

$$\mathcal{L}(\lambda) = \min_{x,\ u} \sum_{i=1}^{n} \sum_{t=1}^{T} f_i(x_t^i, u_{t-1}^i, u_t^i) - \sum_{t=1}^{T} \lambda_t (\sum_{i=1}^{n} x_t^i - d_t)$$

---

[1] To simplify notation, we assume that the penalty term, $\mu_t^{i,s}(u_t^{i,s} - c_k^i)$, is included in the function f_i.

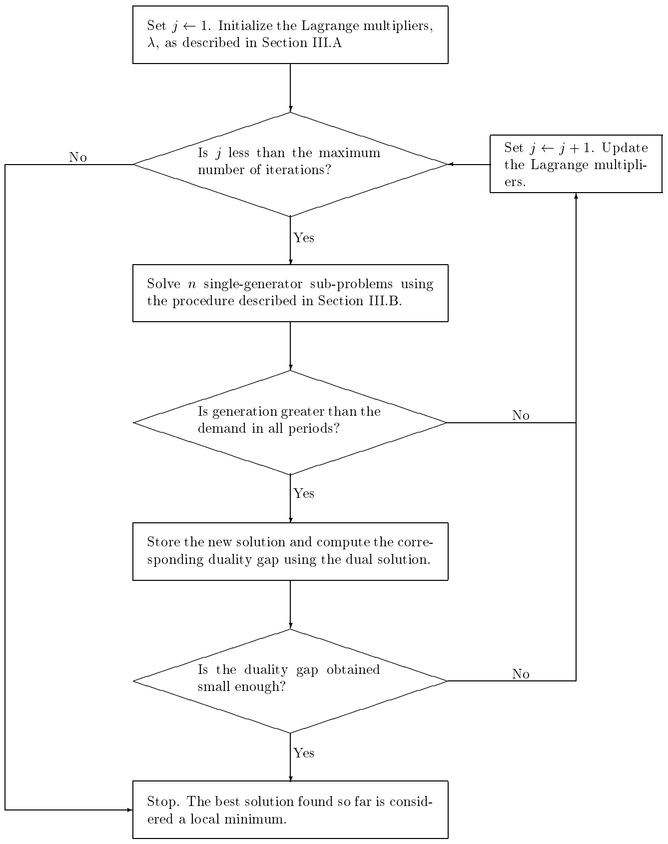

Figure 3: The Lagrangian Relaxation Approach

subject to unit minimum and maximum operating levels, and minimum on-time and off-time constraints. The Lagrangian dual function can be rewritten as

$$\mathcal{L}(\lambda) = \min_{x,u} \sum_{i=1}^{n} \sum_{t=1}^{T} f_i(x_t^i, u_{t-1}^i, u_t^i) - \frac{\lambda_t}{n}(x_t^i - d_t).$$

For a given λ, the value of $\mathcal{L}(\lambda)$ is computed by solving n mixed-integer quadratic programs

$$\phi_i(\lambda) = \min_{x,\ u} \sum_{t=1}^{T} \left(f_i(x_t^i, u_{t-1}^i, u_t^i) - \frac{\lambda_t}{n}(x_t^i - d_t) \right), \tag{19}$$

and adding the resulting values of $\phi_i(\lambda)$ over all generators. Note that each of these programs is independent of the others, which motivates a parallel implementation of this step. The optimization problem in (19) is called the i^{th} single-generator sub-problem since it only depends on the specifications of generator i.

If the primal solution corresponding to a given λ is feasible, then the Lagrange function value is a lower bound on the primal objective function value (weak duality). It follows that the maximum dual objective function value is a lower bound on the primal optimal objective function value. The difference between the optimal value of the original program and the optimal value of the Lagrangian dual problem is called the duality gap. It is expected to be strictly positive since the feasible region of the relaxed problem is not convex.

One can also show that the Lagrange function is concave, hence continuous, in the parameter λ and that it is bounded, which implies that a global optimum can be reached by using an appropriate convex programming method. The two previous remarks are the main attraction in this technique since one can replace a hard primal problem by that of maximizing a concave function. Note that the Lagrange function is not differentiable at all points λ, which complicates the process of maximizing the dual function.

To solve the dual problem, a starting point, λ, is chosen according to some criterion such as described below. Then, the value of the Lagrange function, $\mathcal{L}(\lambda)$, is computed by solving n minimization problems. If the resulting primal solution is feasible, then $\mathcal{L}(\lambda)$ is a lower bound and $\sum_i F_i(\lambda)$ is an upper bound on the optimal value of (18). If the difference between the upper and lower bounds is relatively small, the procedure terminates. Otherwise, the process is repeated by choosing a better Lagrange multiplier. A more detailed description of the steps of this process is provided in the following subsections.

8.1 Initial Lagrange Multipliers

Clearly, the choice of the initial values of the Lagrange multipliers is crucial since it affects the number of iterations, and hence the execution time, of the algorithm. Here, we describe a technique that we found to provide good approximations for λ efficiently.

Since the cost function of each generator, $f_i'(x)$, is a quadratic convex function, one can obtain a starting λ by solving the quadratic program:

$$\begin{array}{lrcl} \min_x & \sum_{i=1}^n \sum_{t=1}^T f_i'(x_t^i) + \frac{S_i}{G_i} x_t^i & & \\ \text{s.t.} & \sum_{i=1}^n x_t^i & \geq & d_t, \\ & 0 \leq x_t^i & \leq & G_i. \end{array} \tag{20}$$

In other words, the minimum on and minimum off time requirements, and the lower bounds on generation capacities, are relaxed. The term $\frac{S_i}{G_i} x_t^i$ takes into consideration the start-up cost of each unit.

The program in (20) is a convex quadratic program that can be solved using any quadratic programming algorithm. Note that the previous quadratic program is separable and can be written as the sum of T quadratic programs, each of which has exactly one constraint:

$$\begin{array}{lrcl} \min_x & \sum_{i=1}^n f_i'(x_t^i) + \frac{S_i}{G_i} x_t^i & & \\ \text{s.t.} & \sum_{i=1}^n x_t^i & \geq & d_t \\ & 0 \leq x_t^i & \leq & G_i. \end{array} \tag{21}$$

The dual variables associated with the demand constraints, $\sum_{i=1}^n x_t^i \geq d_t$, are used as initial Lagrange multipliers for (18).

We can make the program in (21) separable by associating a Lagrange multiplier, $\mu_t \geq 0$, with the demand constraint. Now, we have n different quadratic convex programs, each of which has one bounded decision variable, x_t^i. The solution is found easily by finding the point, $\overline{x}_t^i(\mu_t)$, at which the derivative of the objective function vanishes, and by then choosing the closest feasible point, $x_t^{i*}(\mu_t)$, to $\overline{x}_t^i(\mu_t)$ as an optimal solution. The value of the dual multiplier, μ_t, is then updated. The process is repeated until $\sum_{i=1}^n x_t^{i*}(\mu_t)$ is equal to d_t. The final values of μ_t are used as an approximation for the Lagrange multipliers for the mixed-integer quadratic program in (19).

To improve the efficiency of the procedure, one can build a table that contains the demand values corresponding to different marginal costs, μ_t, at the beginning of the solution process. In other words, this table is generated by assuming certain values for μ_t and finding the corresponding d_t. This approximation was implemented and, surprisingly, the dual objective function value of the initial solution was always within one percent of the optimal dual objective function value.

8.2 Solving the Single-Generator Sub-Problem

We use dynamic programming to solve the mixed-integer quadratic program in (19). At each time period or stage, a generator can be in one of two states: on or off. The formulation needs to take into consideration the start-up and shut-down costs as well as the minimum on-time and minimum off-time constraints. That is, a variable must represent the number of periods spent so far in the current state. Many researchers suggested using $L + l$ nodes at each stage of the dynamic programming graph [30, 29, 40]. No change in the generator state is permitted at nodes where a generator has been on/off for less than L/l periods.

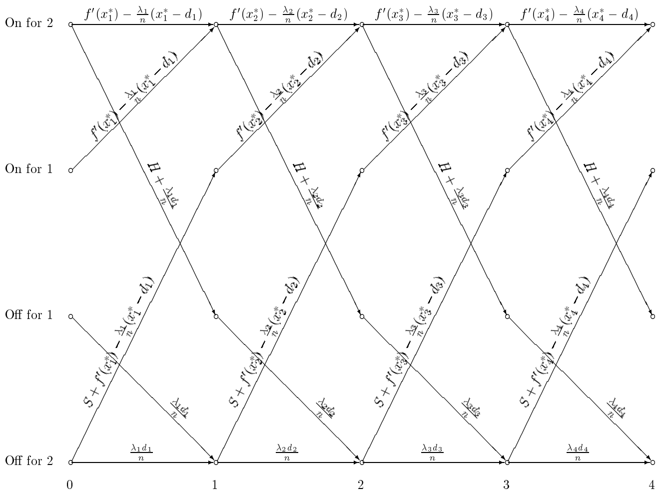

Figure 4: Traditional D.P. Formulation for Single-Generator Sub-Problem

The following observations reduce the number of nodes in the dynamic programming graph substantially when the start-up and shut-down costs, S_i and H_i, are independent of how long a unit has been off or on:

- Since one can only make decisions at those nodes in which a generator has been on/off for at least L/l periods, one needs only to consider these nodes in the dynamic programming formulation.
- Furthermore, the decision set of a generator that has been on/off for longer than L/l is the same as that of a generator that has been on/off for exactly L/l periods.

So, it suffices to consider two nodes at each stage: "on" for at least L periods and "off" for at least l periods. In other words, the number of nodes in the dynamic programming graph is $2T$, which seems to be best in this case.

Each node is represented by two indices. The first index indicates the state of the generator; "on" or "off", and the second one indicates the time or the stage variable. For instance, the node $(\text{off}, 3)$ indicates that at time period 3, the current generator has been off for at least l periods. Hence, this generator can be switched on, that is, move to node $(\text{on}, 3+L)$, at a cost of

$$c(\text{off}, 3) + \sum_{t=4}^{3+L} \min_{x_t^i} \left(f_i'(x_t^i) - \frac{\lambda_t}{n}(x_t^i - d_t) \right) + S_i.$$

On the other hand, we can keep the generator off, that is, move to the node $(\text{off}, 4)$ at a cost of $c(\text{off}, 3) + \lambda_t d_t / n$. The cost at any node, c, represents the minimal cost needed to reach this node in the graph given the initial state at time 0.

Given a Lagrange multiplier, λ, the following algorithm describes the steps required to find an optimal solution for the single generator problem[2].

- Initialization. Let x_t^*, $t = 1, \ldots, T$, be an optimal solution for the convex optimization problem:

  $$\min_{x_t} f'(x_t) - \frac{\lambda_t}{n}(x_t - d_t), \; t = 1, \ldots, T,$$

  such that x_t^* is within the operating range of that generator. Let $c(\text{off}, t) = \infty$ for all $t = 0, \ldots, T + l - 1$, and $c(\text{on}, t) = \infty$ for all $t = 0, \ldots, T + L - 1$. If the initial state is off, then let $c(\text{off}, 0) = 0$; otherwise, let $c(\text{on}, 0) = 0$. Set $t \leftarrow 0$.

- General Step.

$$\begin{aligned}
1.\ c(\text{off}, t+1) &= \min\{c(\text{off}, t+1) c(\text{off}, t) + \tfrac{\lambda_{t+1}}{n} d_{t+1}\}. \\
c(\text{on}, t+L) &= \min\{c(\text{on}, t+L), c(\text{off}, t) \\
&\qquad + \textstyle\sum_{\tau=t+1}^{\min\{t+L,T\}} (f'(x_\tau^*) - \tfrac{\lambda_\tau}{n}(x_\tau^* \\
&\qquad - d_\tau)) + S\}. \\
c(\text{on}, t+1) &= \min\{c(\text{on}, t+1), c(\text{on}, t) + f'(x_{t+1}^*) \\
&\qquad - \tfrac{\lambda_{t+1}}{n}(x_{t+1}^* - d_{t+1})\}. \\
c(\text{off}, t+l) &= \min\{c(\text{off}, t+l), c(\text{on}, t) \\
&\qquad + \textstyle\sum_{\tau=t+1}^{\min\{t+l,T\}} \tfrac{\lambda_\tau}{n} d_\tau + H\}.
\end{aligned}$$

---

[2] To simplify notation, the generator's index i will be skipped throughout the steps of the algorithm.

2. Set $t \leftarrow t + 1$.
3. If $t < T$ then go to Step 1.
4. Check the nodes (off, t), $t = T, \ \ldots, \ T + l - 1$, and (on, t), $t = T, \ \ldots, \ T + L - 1$, to find the optimal value of the objective function corresponding to the given λ.

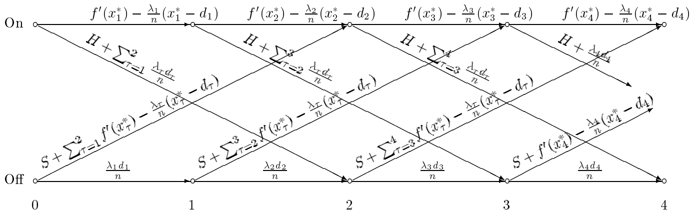

Figure 5: Suggested D.P. Formulation for the Single-Generator Sub-Problem

Note that the number of functional evaluations needed to compute the summation in Step 1 is only one. A predecessor pointer is associated with each node. Whenever the cost, c, of a node changes, the pointer is updated to point to that node from which the transition has the lowest cost. At termination, one can use the predecessor pointers recursively to find an optimal strategy.

8.3 Updating Lagrange Multipliers

If the primal solution corresponding to the current dual solution, λ, is infeasible, or if the primal solution is feasible but the duality gap is not small enough, then a new dual vector must be used so that a better primal feasible solution is obtained. Since the dual objective function is concave, but not everywhere differentiable, it may not have a gradient everywhere. A subgradient technique is necessary to maximize this function. A vector ξ is called a subgradient of $\mathcal{L}(.)$ at $\overline{\lambda}$ if

$$\mathcal{L}(\lambda) \leq \mathcal{L}(\overline{\lambda}) + (\lambda - \overline{\lambda})^T \xi.$$

If the subgradient is unique at a point λ, then it is the gradient at that point. The set of all subgradients at λ is called the subdifferential, $\partial \mathcal{L}(\lambda)$, and is a closed convex set. A necessary and sufficient condition for optimality in subgradient optimization is $0 \in \partial \lambda$.

We used the subgradient optimization algorithm to generate a sequence of points using the rule

$$\lambda^{l+1} = \lambda^l + \alpha_l \xi^l,$$

where ξ^l is any subgradient of $\mathcal{L}(.)$ at λ^l [38]. The step size, α_l, has to be chosen carefully to achieve good performance by the algorithm. Poljak [32] has shown that the necessary and sufficient conditions to guarantee the convergence of the sequence λ^l to an optimal solution is

$$\alpha_l \|\xi^l\| \to 0 \text{ and } \sum_l \alpha_l \|\xi^l\| \to \infty, \tag{22}$$

where $\|.\|$ denotes the Euclidean norm. In fact, a geometric convergence rate can be achieved if we choose

$$\alpha_l = \frac{\mathcal{L}^* - \mathcal{L}(\lambda^l)}{\|\xi^l\|^2},$$

where $\mathcal{L}^*$ is the optimal value of the Lagrange function. A problem here is that $\mathcal{L}^*$ is not known in advance. We adopt a simple updating rule that satisfies (22) but has no known theoretical convergence rate

$$\alpha_l = \frac{1}{a + bl},$$

where a and b are constants chosen according to the given data. We also impose an upper bound on the step size to maintain dual feasibility.

Recall that our problem is to maximize

$$\mathcal{L}(\lambda^l) = \min_{x,\ u} \sum_{i=1}^{n} \sum_{t=1}^{T} f_i(x_t^i, u_{t-1}^i, u_t^i) - \sum_{t=1}^{T} \lambda_t^l \Big(\sum_{i=1}^{n} x_t^i - d_t\Big)$$

subject to $\lambda^l \geq 0$. A subgradient of the function $\mathcal{L}(\lambda^l)$ with respect to λ^l is given by

$$\nabla_t \mathcal{L}(\lambda^l) = d_t - \sum_{i=1}^{n} x_t^i, \ t = 1, \ \ldots, \ T.$$

If $\nabla_t \mathcal{L}(\lambda^l)$ is not equal to zero, then it is an ascent direction. The value of $\mathcal{L}(\lambda^l)$ will increase by moving to $\lambda_t^l + \alpha_l(d_t - \sum_i x_t^i)$, $\alpha_l \geq 0$. Since we want to maintain dual feasibility, that is, $\lambda_l \geq 0$, the maximum step length is

$$\alpha_{\max} = \frac{\lambda_t^l}{\sum_{i=1}^{n} x_t^i - d_t} \tag{23}$$

when $\sum_{i=1}^{n} x_t^i \geq d_t$. In the case where the primal solution is infeasible, $\sum_{i=1}^{n} x_t^i \leq d_t$, there is no upper bound on the value of α_l, but we impose a maximum step length in the implementation. The previous points can be summarized as follows:

- If $\lambda_t^l = 0$ and $\sum_i x_t^i \geq d_t$, then set $\lambda_t^{l+1} \leftarrow \lambda_t^l$.
- If $\lambda_t^l > 0$ and $\sum_i x_t^i \geq d_t$, then set $\lambda_t^{l+1} \leftarrow \lambda_t^l + \alpha_l(d_t - \sum_i x_t^i)$, where α_l does not exceed $\frac{\lambda_t^l}{\sum_i x_t^i - d_t}$.
- If $\sum_i x_t^i \leq d_t$, then set $\lambda_t^{l+1} \leftarrow \lambda_t^l + \alpha_l(d_t - \sum_i x_t^i)$.

The previous procedure is efficient and simple to implement.

9 Scenario Generation

In order to obtain a policy that is of practical use, we need to provide the progressive hedging algorithm with scenarios that truly reflect all possible future demands. Furthermore, the probabilities assigned to these scenarios must be calculated carefully. Clearly, this is not an easy task.

More research is needed in this area to develop better understanding of the demand randomness and the related factors. One thing which must be considered in regards to scenario generation is that the more scenarios which are created, the better the hedging policy of the algorithm. On the other hand, the execution time of the algorithm grows rapidly as the number of scenarios included increases and as their demands are more diverse.

We have developed a neural network model to identify these scenarios based on the fuzzy neural network approach with back propogation as in Kwan [25]). The basic approach is to train the network to identify weights on historical patterns so that both point forecast and scenario sets can be identified. Using a variety of heuristic techniques, we have been able to produce 10-day forecasts for the Michigan power network with approximately 1% accuracy for peakload weekday forecasts. This level of accuracy allows significant improvement in the effectiveness of our unit commitment models.

Among the factors that randomly affect the load on the system are generator failures and unexpected or unusual demands. One should also consider the available electricity that can be purchased from other suppliers and the sales of electricity to other regions with emergency needs. Here is a closer look at some of these factors and the suggested methods of treating them in our model.

9.1 Generator Failures

A given generator may be unavailable for use at any time due to unforeseen circumstances, such as mechanical problems, unscheduled maintenance, etc. The current strategy of most producers is to maintain adequate operating reserves for these types of situations. At any given time, some of these operating reserves are available from sources such as a pump-storage plant (as in our example for the Michigan power companies, Detroit Edison and Consumers Power, who participate in the Michigan Power Coordination Center).

We model this problem by creating a scenario which has demand increases equal to the unavailable generator's capacity. These demand increases occur during the periods in which the generator is expected to be down. The scenario is then assigned a weight equal to the probability that this generator will be unavailable. Different scenarios can be created for different generator failures.

Another way to model this problem is to approximate the generating capacity loss over a certain period of time. The advantage of this technique is its independence from the individual generators. For instance, Michigan Electric Power that covers all possible increases. We use this technique to model uncertainty in the numerical example of Section 10.

9.2 Inaccurate Forecast

Experienced schedulers forecast the load for each hour of the week on Sunday and, if necessary, update this forecast on each shift. To estimate the load on the system, a scheduler uses the data of the same week from previous years, the data of days with similar weather conditions, and his/her personal experience. The resulting load forecast is then used as the input for the unit commitment program which determines an optimal strategy.

Here again, we created several scenarios with high demand increases or decreases. The unusual changes in the demand are assigned probabilities according to the likelihood of their occurrence. This approach makes it easy for a scheduler to include all the relevant data of a certain week in the model after assigning a suitable weight to each of them.

9.3 Other Factors

Electricity can be bought from or sold to other electric companies. The price per Megawatt-Hour varies depending on the time of the day, the contracts governing the trade between these companies, the amount of electricity needed, the urgency of the electricity need, etc. All the resulting changes in the load on the system can be modeled as possible scenarios. Each of these scenarios has an additional generator with an appropriate cost function corresponding to the price of the available electricity. The minimum and maximum generating capacities of such a generator are set equal to the total amount of electricity that can be purchased in each period. Each generator of this type is a must-run unit in its corresponding scenario and is assumed to be unavailable in all other scenarios. The only difficulty in applying this approach is that the probability assigned to each scenario needs to be estimated accurately. This estimation can be done easily if there is a large database that will permit good approximation of these probabilities.

10 Computational Experience

The progressive hedging technique was implemented using the C language on an IBM RS 6000. We also developed a parallel version of this code to be used on a KSR super computer and an IBM SP/1. All of our results in this section are computed on an IBM RS 6000/320H. The code, which contains 4000 lines of C, is written in such a way that it makes full use of the RISC architecture. A unit commitment problem with 2400 binary variables and 2400 continuous variables is solved in less than one second to within 0.1% of the optimal solution. A typical weekly, stochastic unit commitment problem with 10 scenarios and a 168-hour planning horizon requires approximately one hundred progressive hedging iterations, and six hours of CPU time. For our parallel results, we found speedups of 26 on 32 processors. We believe that similar almost linear speedups in processors would allow solutions in a minute or less.

We used the same data supplied by the Michigan Electric Power Coordination Center. The system has a pump-storage hydro plant with six units and more than one hundred thermal units. Water is pumped into the reservoir during low-demand periods so that it can be used during peak-demand periods. For this study, the power needed to pump the water into the storage facility is assumed to be proportional to the change in the water level.

The thermal units are divided into four different categories. The first consists of the must-run units which are assumed to be online everyday throughout the year. These units have relatively lower operating costs compared with the others. The second category consists of the units that are unavailable due to maintenance, mechanical failures, or other factors. Cyclers form another category. These units must be scheduled for use in advance, since they require a long time to get online. The last category consists of peakers. The other scenarios were generated from scenario one by adding 400 Megawatt-

Hours with probability one-third at the beginning of Tuesday, Wednesday, Thursday, and Friday. This increase in the demand takes into account the possible failure of some thermal units. Table 1 summarizes the demand increases of different scenarios relative to the expected demand. The scenario tree and different scenario bundles can be seen in Figure 6.

Table 1: Demand Increases for Different Scenarios

Scen.	Prob.	Mon.	Tue.	Wed.	Thu.	Fri.
1	19.05%	0	0	0	0	0
2	9.52%	0	0	0	0	400
3	9.52%	0	0	0	400	400
4	4.76%	0	0	0	400	800
5	0.95%	0	0	0	+20%	+20%
6	9.52%	0	0	400	400	400
7	4.76%	0	0	400	400	800
8	4.76%	0	0	400	800	800
9	1.90%	0	0	400	800	1200
10	0.95%	0	0	400	800	+15%
11	0.95%	0	0	+20%	+20%	+20%
12	9.52%	0	400	400	400	400
13	4.76%	0	400	400	400	800
14	0.95%	0	400	400	400	+20%
15	4.76%	0	400	400	800	800
16	1.90%	0	400	400	800	1200
17	0.95%	0	400	400	+15%	+15%
18	4.76%	0	400	800	800	800
19	1.90%	0	400	800	800	1200
20	1.90%	0	400	800	1200	1200
21	0.95%	0	400	800	1200	1600
22	0.95%	0	400	+20%	+20%	+20%

Table 2 shows the costs of applying the optimal strategy for a given scenario to each of the other scenarios. Obviously, the lowest cost for each scenario occurs when its own optimal strategy is used. The expected value of using each optimal strategy is computed and given in the column $E(x)$. The costs corresponding to applying the progressive hedging policy to all scenarios are provided in the last row of Table 2. The difference between the expected value of each optimal policy and that of the progressive hedging

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	$E(x)$	Diff.	%
Pr	19.05%	9.52%	9.52%	4.76%	0.95%	9.52%	4.76%	4.76%	1.90%	0.95%	0.95%	9.52%	4.76%	0.95%	4.76%	1.90%	0.95%	4.76%	1.90%	1.90%	0.95%	0.95%			
1	19247	19416	19583	19823	22812	19748	19989	20236	20584	22245	24573	19915	20156	23120	20403	20742	23727	20647	20986	21344	21689	29673	20124	145.12	0.73%
2	19284	19416	19613	19788	22268	19778	19953	20200	20409	21457	24032	19945	20120	22263	20367	20577	22938	20611	20821	21190	21426	28819	20075	96.87	0.48%
3	19270	19438	19583	19787	22242	19763	19951	20134	20438	21903	24003	19930	20117	22840	20301	20605	23153	20545	20848	21124	21482	29078	20072	93.63	0.47%
4	19300	19462	19626	19786	21939	19791	19963	20150	20348	21341	23700	19959	20131	22172	20317	20515	22586	20561	20759	21066	21312	28408	20056	77.01	0.39%
5	19429	19589	19752	19922	21324	19918	20088	20256	20450	21156	23085	20086	20256	21638	20424	20618	21928	20668	20862	21056	21264	26444	20123	144.57	0.72%
6	19280	19448	19608	19797	22270	19747	19955	20144	20442	21933	23746	19938	20122	22871	20312	20603	23186	20493	20803	21085	21439	28815	20071	92.83	0.46%
7	19311	19472	19636	19808	21960	19802	19950	20157	20357	21347	23441	19968	20140	22182	20324	20524	22604	20513	20708	21022	21266	28139	20055	76.10	0.38%
8	19343	19504	19668	19840	21599	19833	20006	20134	20370	21362	23087	20001	20173	22214	20343	20538	22290	20527	20723	20925	21162	27417	20062	83.33	0.42%
9	19344	19505	19669	19841	21594	19834	20007	20176	20348	21354	23076	20002	20174	22197	20344	20539	22294	20527	20723	20925	21159	27394	20062	83.53	0.42%
10	19395	19555	19719	19889	21471	19884	20053	20224	20419	21119	22947	20053	20223	21588	20392	20585	22056	20575	20769	20966	21171	26771	20090	111.04	0.56%
11	19519	19678	19841	20011	21413	20006	20176	20344	20537	21242	22347	20174	20344	21727	20512	20705	22010	20680	20873	21067	21276	24838	20180	201.68	1.01%
12	19291	19459	19618	19809	22293	19781	19967	20157	20455	21952	23774	19914	20131	22887	20318	20608	23209	20506	20806	21091	21453	28846	20082	103.86	0.52%
13	19319	19480	19644	19817	22003	19810	19981	20166	20375	21419	23484	19975	20116	22288	20334	20535	22679	20516	20725	21023	21288	28246	20066	88.00	0.44%
14	19396	19555	19719	19889	21840	19883	20052	20236	20429	21134	23316	20051	20219	21578	20404	20595	22370	20586	20779	21075	21279	27449	20111	132.44	0.66%
15	19349	19510	19674	19846	21658	19839	20011	20182	20385	21436	23147	20006	20178	22319	20301	20552	22387	20532	20736	20939	21198	27571	20073	94.92	0.48%
16	19354	19515	19678	19850	21621	19843	20017	20186	20381	21365	23110	20011	20182	22206	20353	20514	22317	20536	20731	20934	21173	27459	20072	93.94	0.47%
17	19454	19613	19777	19946	21348	19943	20113	20279	20472	21175	22829	20109	20278	21644	20447	20639	21927	20630	20823	21018	21225	26162	20135	156.58	0.78%
18	19379	19540	19703	19875	21688	19869	20040	20213	20414	21465	22829	20035	20207	22348	20378	20582	22416	20491	20752	20951	21207	26930	20092	113.26	0.57%
19	19383	19544	19707	19879	21651	19873	20046	20215	20411	21392	22791	20040	20212	22235	20382	20576	22340	20553	20708	20946	21182	26818	20091	112.05	0.56%
20	19387	19548	19711	19883	21635	19877	20050	20219	20413	21401	22757	20044	20216	22239	20386	20580	22324	20557	20751	20923	21181	26718	20093	114.17	0.57%
21	19420	19580	19744	19914	21497	19909	20079	20249	20442	21160	22637	20077	20246	21664	20417	20609	22093	20587	20780	20976	21156	26193	20105	126.20	0.63%
22	19555	19714	19877	20046	21446	20042	20212	20378	20571	21272	22380	20209	20378	21735	20546	20740	22039	20714	20906	21099	21309	24813	20214	235.95	1.18%
Policy	19247	19446	19598	19798	21324	19769	19972	20176	20371	21119	22347	19944	20150	21578	20348	20547	21942	20549	20747	20945	21193	24813	19978	0.00	0.00%

Table 2: The Cost in \$1000 of Applying the Policy of Each Scenario to All Other Scenarios

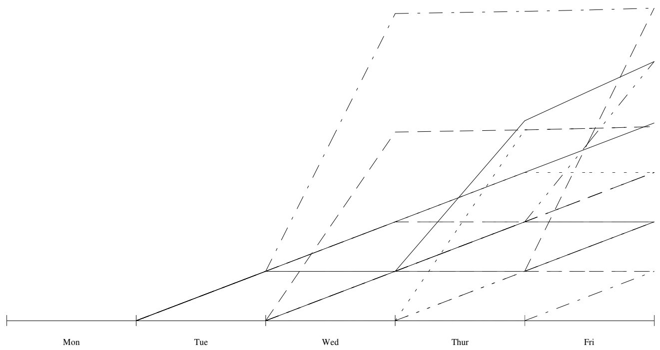

Figure 6: Scenario Tree for the Example

represents the saving incurred by using progressive hedging. The expected cost corresponding to the progressive hedging strategy, $19,978,000, is lower than that of the deterministic strategy, $20,124,000, yielding a saving of $146,000 for this week. As a matter of fact, in the long run, the cost of applying the progressive hedging strategy must be less than the expected cost of applying any other strategy to all scenarios.

10.1 Inaccurate Forecast Example

We created sixteen different scenarios using the demand data provided by Detroit Edison and Consumers Power for the months of August in the years 1990, 1991, 1992, and 1993. The four weeks of August in each of the four years are assumed to be probabilistically independent with identical demand distribution. Hence, each week is considered a possible scenario with a weight of $\frac{1}{16}$. To create scenario bundles, we assumed that the demand on Monday, Tuesday, and Wednesday is known in advance and is equal to the average demand over these days in the given sixteen scenarios. On Thursday morning, there are four branches. Each of them is taken as the average of the demand on Thursdays in August in each of the four years. Then, each of Thursday's scenarios splits on Friday morning; the resulting sixteen branches have the same demand on Friday, Saturday, and Sunday as that of the sixteen observed scenarios. Figure 7 shows the resulting scenario tree.

We solved the problem with the Figure 7 scenario tree for each of the individual scenarios to obtain its optimal policy, which is then applied to other scenarios to compute the expected operating cost. We also solved the problem using the expected demand. The results are shown in the "Deterministic" row of Table 3. Progressive hedging is then used to minimize the expected cost over all scenarios. The results are shown in the last row of the same table. As expected, the hedging policy performs better than any other policy. It saves on the average $120,000 over the deterministic policy.

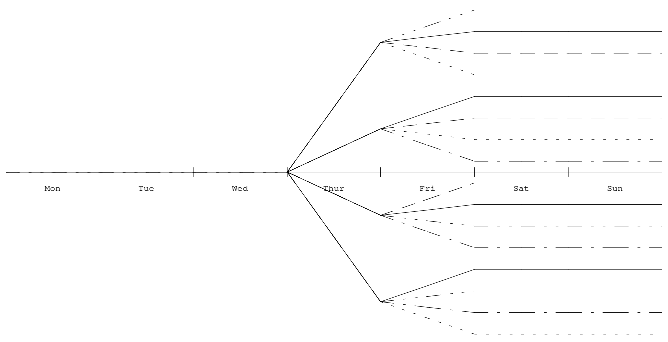

Figure 7: Scenario Tree

11 Overall Impact

We introduced a stochastic model for the unit commitment problem. The uncertainty in the demand is modeled by choosing a set of possible scenarios. Each scenario is assigned a probability that reflects its likelihood of occurrence. The model was solved using a Lagrangian relaxation type of technique, namely, progressive hedging. To solve the deterministic model efficiently, we had to improve the existing methods for solving the unit commitment problem by selecting good starting values for the marginal costs and improving the single-generator sub-problem dynamic programming formulation. Both the stochastic model and the deterministic model can be implemented in parallel to reduce the computation time. The numerical results obtained indicate that using the stochastic model can reduce the cost of an electrical power system significantly.

The unique aspect of this work is our incorporating the logical information level into the system state, building real-time response to that information level, and tying that response into more global information of the planning level schedule. We construct an initial response function that maps key information factors (including perhaps logical statements) into real-time generation allocation decisions. This response function determines from our off-line

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	$E(x)$	Diff.	%
Pr	6.25%	6.25%	6.25%	6.25%	6.25%	6.25%	6.25%	6.25%	6.25%	6.25%	6.25%	6.25%	6.25%	6.25%	6.25%	6.25%			
1	20304	20529	20492	20003	20335	19893	19639	24138	21293	21162	20344	21695	21330	20861	19946	25659	21102	199.05	0.95%
2	20332	20496	20517	20027	20359	19919	19664	23672	21244	21064	20369	21551	21354	20885	19973	25134	21035	132.40	0.63%
3	20304	20566	20489	19999	20334	19890	19633	24234	21325	21194	20340	21759	21329	20858	19942	25757	21122	219.49	1.05%
4	20318	20580	20503	19999	20348	19903	19647	24225	21323	21192	20345	21740	21327	20856	19939	25724	21124	221.50	1.06%
5	20311	20534	20496	20008	20334	19896	19644	24142	21286	21155	20336	21687	21311	20842	19927	25639	21097	194.54	0.93%
6	20321	20583	20506	20017	20351	19890	19649	24251	21324	21193	20339	21759	21307	20837	19920	25736	21125	222.38	1.06%
7	20310	20571	20494	20005	20339	19894	19632	24240	21316	21185	20331	21751	21302	20832	19916	25729	21116	213.29	1.02%
8	20461	20621	20643	20156	20485	20047	19796	23150	21321	21150	20482	21487	21446	20981	20075	24107	21026	123.07	0.59%
9	20379	20540	20565	20073	20406	19965	19710	23555	21240	21071	20392	21480	21330	20864	19962	24824	21022	119.60	0.57%
10	20404	20565	20590	20098	20432	19990	19736	23569	21266	21063	20418	21508	21356	20889	19989	24814	21045	142.50	0.68%
11	20340	20599	20525	20036	20370	19925	19668	24270	21337	21207	20330	21772	21295	20826	19909	25725	21135	232.25	1.11%
12	20413	20572	20599	20107	20440	19999	19745	23443	21271	21103	20427	21433	21364	20907	19988	24329	21009	106.24	0.51%
13	20344	20597	20529	20034	20373	19930	19676	24228	21326	21196	20356	21732	21296	20828	19913	25683	21127	224.93	1.08%
14	20342	20595	20527	20038	20371	19928	19673	24228	21323	21194	20354	21730	21295	20826	19911	25681	21126	223.57	1.07%
15	20340	20600	20526	20036	20370	19926	19668	24270	21338	21207	20353	21773	21296	20827	19908	25726	21135	232.72	1.11%
16	20472	20646	20666	20177	20523	20068	19816	23347	21336	21163	20496	21498	21432	20976	20056	23970	21042	139.43	0.67%
Deter	20381	20563	20557	20082	20419	19987	19746	23090	21241	21187	20404	21426	21330	20885	20087	24977	21023	120.09	0.57%
Policy	20306	20497	20489	20000	20341	19894	19642	23162	21240	21075	20352	21433	21296	20826	19910	23977	20903	0.00	0.00%

Table 3: The Cost in $1000 of Applying the Policy of Each Scenario to All Other Scenarios

optimization routines in conjunction with intelligent searches to determine whether fundamental parameters may have changed.

In comparison to current systems, the on-line control does not simply follow a hierarchy of unit allocations in a myopic procedure. It involves a more comprehensive set of system wide responses depending on current information. These responses are built over time by searching off-line to determine the possible effects of different decisions for similar information states. Actual experience is enhanced by simulated conditions that are also consistent with given infomation states. As the system evolves, progressively improved decision sets form the basis of secure and efficient real-time generation.

These searches of possible on-line responses also feed into a longer term unit commitment planning part of the system. In this way, both on-line and off-line decision levels are simultaneously improved. The result is a system that is most robust to changing conditions and more efficient in its allocation.

12 Acknowledgments

We would like to thank the staff of the Michigan Power Coordination Center for providing the data used in testing our strategy. Special thanks go to Gregory J. Ostrowski for his helpful comments regarding the model.

13 References

Bibliography

[1] K. Aoki, M. Itoh, T. Satoh, K. Nara, and M. Kanezashi. Optimal long-term unit commitment in large scale systems including fuel constrained thermal and pumped-storage hydro. *IEEE Transactions on Power Systems*, 4(3):1065–1073, August 1989.

[2] M. Avriel, G.B. Dantzig, and P.W. Glynn. Decomposition and Parallel Processing for Large-Scale Electric Power Systems Planning under Uncertainty. In: G. Dantzig, P. Glynn, *Resource Planning under Uncertainty for Electric Power Sytems*, NSF, 1989.

[3] C.H. Bannister and R.J. Kaye. A Rapid Method for Optimization of Linear Systems with Storage, *Operation Research* 39: 220-232, 1991.

[4] J. Bean, J. Birge, J. Mittenthal, and C. Noon. Matchup scheduling with multiple resources, release dates, and disruptions. *Operations Research* 39: 470-483, 1991.

[5] D. P. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods*, chapter 5.6, page 368. Academic Press, 1982.

[6] D. P. Bertsekas, G. S. Lauer, N. R. Sandell, and T. A. Posbergh. Optimal short-term scheduling of large-scale power systems. *IEEE Transactions on Automatic Control*, 28(1):1–11, January 1983.

[7] J.R. Birge. Decomposition and partitioning methods for multi–stage stochastic linear programs. *Operations Research* 33:. 989-1007, 1985.

[8] J.R. Birge. Aggregation Bounds in Stochastic Linear Programming. *Mathematical Programming* 31: 25-41, 1985.

[9] J.R. Birge. Exhaustible Resource Models with uncertain returns from Exploration Investment. In: *Numerical Techniques for Stochastic Optimization*, ed. Y. Ermoliev & R. J.-B. Wets, pp. 481- 488, 1988.

[10] J.R. Birge. Multistage stochastic planning models using piecewise linear response functions. In: G. Dantzig, P. Glynn, *Resource Planning under Uncertainty for Electric Power Sytems*, NSF, 1989.

[11] J.R. Birge and M.A.H. Dempster. Optimality conditions for match-up strategies in stochastic scheduling and related dynamic stochastic optimization problems. Technical Report, IOE Department, University of Michigan, Ann Arbor, MI, 1992.

[12] J.R. Birge and R. J–B. Wets. Designing approximation schemes for stochastic optimization problems, in particular, for stochastic programs with recourse. *Math. Progr. Study* 27: 54-102, 1986.

[13] J.A. Bloom. Solving an Electricity Generating Capacity Expansion Planning Problem by Generalized Benders' Decomposition. *Operations Research* 31: 84–100, 1983.

[14] A.B. Borison, P.A. Morris, and S. S. Oren. A State-of-the-Workd Decomposition Approach to Dynamics and Uncertainty in Electric Utility Generation Expansion Planning. *Operations Research* 32: 1052–1068, 1984.

[15] B.H. Chowdhury and S. Rahman. A Review of Recent Advances in Economic Dispatch, *IEEE Trans. on Power Systems* 5, No.4: 1248–1257, 1990.

[16] A. I. Cohen, and V. R. Sherkat. Optimization-Based Methods for Operations Scheduling. *IEEE Trans. on Power Systems*, No.12, 1987.

[17] G. Dantzig and P. Glynn. Parallel processors for planning under uncertainty. *Annals of Operations Research* 22: 1-21, 1990.

[18] I. Deák, J. Hoffer, J. Mayer, A. Németh, B. Potecz, A. Prékopa, and B. Strazicky. Recent Advances Concerning the Problem of Optimal Daily Scheduling of Electricity Production in Hungary. In: *IFAC Congress, Budapest, Vol. I*, pp. 95-99, 1984.

[19] East Central Area Reliability Group. Reliability criterion, 1981. Document #1.

[20] S. Granville and M.V.F. Pereira. Analysis of the Linearized Power Flow Model in Benders Decomposition. Technical Report SOL 85-4, Systems Optimization Laboratory, Stanford University, Stanford, CA, 1985.

[21] N. Gröwe and W. Römisch. A Stochastic Programming Model for Optimal Power Dispatch: Stability and Numerical Treatment. Preprint Nr. 266, Section Mathematik, Humboldt-Universität zu Berlin, Berlin, Germany, 1990.

[22] H. H. Happ. The Inter-Area Matrix: A Tie Flow Model for Power Pools. *IEEE Trans. on PAS* 90, No.1: 36–45, 1971.

[23] L. K. Kirchmayer. *Economic Control of Interconnected Systems*, John Wiley & Sons, Inc., 1959.

[24] S. Kirkpatrick, C.D. Gelatt, and M.P. Vecchi. Optimization by Simulated Annealing, *Science* 200: 671-680, 1983.

[25] H. K. Kwan. A fuzzy neural network and its application to pattern regocnition. *IEEE Transactions on Fuzzy Systems* 2, No. 3, 1994.

[26] K. D. Le, R. R. Jackups, J. Feinstein, and J. S. Griffith Operational Aspects of Generating Cycling. *IEEE Transactions on Power Systems*, 5(4):213–217, November 1991.

[27] F. V. Louveaux. Multistage Stochastic Programs with Block Separable Recourse. *Mathematical Programming Study* 28: 48-62, 1986.

[28] F. V. Louveaux. A solution method for multistage stochastic programs with recourse with application to an energy investiment problem. *Operations Research* 28: 889-902, 1980.

[29] A. Merlin and P. Sandrin. A new method for unit commitment at Electricité de France. *IEEE Transactions on Power Apparatus and Systems*, PAS-102(5):1218–1225, May 1983.

[30] J. A. Muckstadt and S. A. Koenig. An application of Lagrangian relaxation to scheduling in power-generation systems. *Operations Research*, 25(3):387–403, May-June 1977.

[31] M.V.F. Pereira and L.M.V.G. Pinto. Stochastic Optimization of a Multireservoir Hydroelectric System: a Decomposition Approach. *Water Resources Research* 21: 779-792, 1985.

[32] B. T. Poljak. A general method for solving extremal problems. *Soviet Mathematics Doklady*, 8:593–597, 1967.

[33] B. Prince, K.-D. Le, R. H. Vierra, G. D. Nagel and R. T. Jenkins. Problems Associated with Unit commitment in Uncertainty. *IEEE Trans. on PAS* 104, No. 8: 2072–2078, 1985.

[34] R. T. Rockafellar and R. J.-B. Wets. Scenarios and policy aggregation in optimization under uncertainty. *Mathematics of Operations Research*, 16(1):119–147, 1991.

[35] Y. Roth and R. Jain. Context-based caching for control and simulation in mobile platforms. In *Proceedings of SimTec'91, Simulation Technology International Conference*, pages 559 – 564. The Society for Computer Simulation (SCS), October 1991.

[36] Y. Roth and R. Jain. Verification versus discovery in vision based systems. Technical Report CSE-TR-110-91, Computer Science and Engineering Division The University of Michigan, Ann Arbor, MI 48109-2122, Room 3402 -EECS Building, September 1991.

[37] S. Ruzić and N. Rajaković. A new approach for solving extended unit commitment problem. *IEEE Transactions on Power Systems*, 6(1):269–277, February 1991.

[38] J. F. Shapiro. *Mathematical Programming: Structures and Algorithms.* John Wiley & Sons, Inc., 1979.

[39] L. Stougie. Design and Analysis of Algorithms for Stochastic Integer Programming. Ph.D. Thesis, Center For Mathematics and Computer Science, Amsterdam, 1985.

[40] F. Zhuang and F. D. Galiana. Towards a more rigorous and practical unit commitment by Lagrangian relaxation. *IEEE Transactions on Power Systems*, 3(2):763–773, May 1988.