Wollenberg, Bruce F. “Power System Operation and Control”
The Electric Power Engineering Handbook
Ed. L.L. Grigsby
Boca Raton: CRC Press LLC, 2001

12

Power System Operation and Control

Bruce F. Wollenberg
University of Minnesota

12

Power System Operation and Control

K. Neil Stanton
Stanton Associates

Jay C. Giri
ALSTOM ESCA Corporation

Anjan Bose
Washington State University

Charles W. Richter, Jr.
ALSTOM ESCA Corporation

Danny Julian
ABB Power T&D Company

M. E. El-Hawary
Dalhousie University

Nouredine Hadjsaid
Institut National Polytechnique de Grenoble (INPG)

12.1 Energy Management

K. Neil Stanton, Jay C. Giri, and Anjan Bose

Energy management is the process of monitoring, coordinating, and controlling the generation, transmission, and distribution of electrical energy. The physical plant to be managed includes generating plants that produce energy fed through transformers to the high-voltage transmission network (grid), interconnecting generating plants, and load centers. Transmission lines terminate at substations that perform switching, voltage transformation, measurement, and control. Substations at load centers transform to subtransmission and distribution levels. These lower-voltage circuits typically operate radially, i.e., no normally closed paths between substations through subtransmission or distribution circuits. (Underground cable networks in large cities are an exception.)

Since transmission systems provide negligible energy storage, supply and demand must be balanced by either generation or load. Production is controlled by turbine governors at generating plants, and automatic generation control is performed by control center computers remote from generating plants. Load management, sometimes called demand-side management, extends remote supervision and control to subtransmission and distribution circuits, including control of residential, commercial, and industrial loads.

FIGURE 12.1 Manitoba Hydro Control Center in Winnipeg, Manitoba, Canada. (Photo used with permission of ALSTOM ESCA Corporation.)

Events such as lightning strikes, short circuits, equipment failure, or accidents may cause a system fault. Protective relays actuate rapid, local control through operation of circuit breakers before operators can respond. The goal is to maximize safety, minimize damage, and continue to supply load with the least inconvenience to customers. Data acquisition provides operators and computer control systems with status and measurement information needed to supervise overall operations. Security control analyzes the consequences of faults to establish operating conditions that are both robust and economical.

Energy management is performed at control centers (see Fig. 12.1), typically called system control centers, by computer systems called *energy management systems* (EMS). Data acquisition and remote control is performed by computer systems called *supervisory control and data acquisition* (SCADA) systems. These latter systems may be installed at a variety of sites including system control centers. An EMS typically includes a SCADA "front-end" through which it communicates with generating plants, substations, and other remote devices.

Figure 12.2 illustrates the applications layer of modern EMS as well as the underlying layers on which it is built: the operating system, a database manager, and a utilities/services layer.

Power System Data Acquisition and Control

A SCADA system consists of a master station that communicates with remote terminal units (RTUs) for the purpose of allowing operators to observe and control physical plants. Generating plants and transmission substations certainly justify RTUs, and their installation is becoming more common in distribution substations as costs decrease. RTUs transmit device status and measurements to, and receive control commands and setpoint data from, the master station. Communication is generally via dedicated circuits operating in the range of 600 to 4800 bits/s with the RTU responding to periodic requests initiated from the master station (polling) every 2 to 10 s, depending on the criticality of the data.

The traditional functions of SCADA systems are summarized:

- Data acquisition: Provides telemetered measurements and status information to operator.
- Supervisory control: Allows operator to remotely control devices, e.g., open and close circuit breakers. A "select before operate" procedure is used for greater safety.
- Tagging: Identifies a device as subject to specific operating restrictions and prevents unauthorized operation.

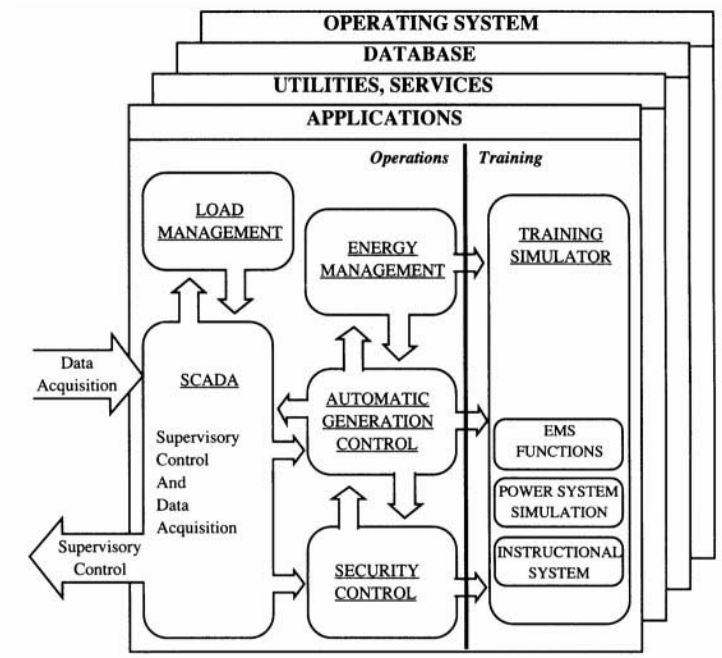


FIGURE 12.2 Layers of a modern EMS.

- Alarms: Inform operator of unplanned events and undesirable operating conditions. Alarms are sorted by criticality, area of responsibility, and chronology. Acknowledgment may be required.
- Logging: Logs all operator entry, all alarms, and selected information.
- Load shed: Provides both automatic and operator-initiated tripping of load in response to system emergencies.
- Trending: Plots measurements on selected time scales.

Since the master station is critical to power system operations, its functions are generally distributed among several computer systems depending on specific design. A dual computer system configured in primary and standby modes is most common. SCADA functions are listed below without stating which computer has specific responsibility.

- Manage communication circuit configuration
- Downline load RTU files
- Maintain scan tables and perform polling
- Check and correct message errors
- Convert to engineering units
- Detect status and measurement changes
- Monitor abnormal and out-of-limit conditions
- Log and time-tag sequence of events
- Detect and annunciate alarms
- Respond to operator requests to:
  - Display information
  - Enter data
  - Execute control action
  - Acknowledge alarms

- Transmit control action to RTUs
- Inhibit unauthorized actions
- Maintain historical files
- Log events and prepare reports
- Perform load shedding

Automatic Generation Control

Automatic generation control (AGC) consists of two major and several minor functions that operate on-line in realtime to adjust the generation against load at minimum cost. The major functions are load frequency control and economic dispatch, each of which is described below. The minor functions are reserve monitoring, which assures enough reserve on the system; interchange scheduling, which initiates and completes scheduled interchanges; and other similar monitoring and recording functions.

Load Frequency Control

Load frequency control (LFC) has to achieve three primary objectives, which are stated below in priority order:

1. To maintain frequency at the scheduled value
2. To maintain net power interchanges with neighboring control areas at the scheduled values
3. To maintain power allocation among units at economically desired values

The first and second objectives are met by monitoring an error signal, called *area control error* (ACE), which is a combination of net interchange error and frequency error and represents the power imbalance between generation and load at any instant. This ACE must be filtered or smoothed such that excessive and random changes in ACE are not translated into control action. Since these excessive changes are different for different systems, the filter parameters have to be tuned specifically for each control area. The filtered ACE is then used to obtain the proportional plus integral control signal. This control signal is modified by limiters, deadbands, and gain constants that are tuned to the particular system. This control signal is then divided among the generating units under control by using participation factors to obtain *unit control errors* (UCE).

These participation factors may be proportional to the inverse of the second derivative of the cost of unit generation so that the units would be loaded according to their costs, thus meeting the third objective. However, cost may not be the only consideration because the different units may have different response rates and it may be necessary to move the faster generators more to obtain an acceptable response. The UCEs are then sent to the various units under control and the generating units monitored to see that the corrections take place. This control action is repeated every 2 to 6 s.

In spite of the integral control, errors in frequency and net interchange do tend to accumulate over time. These time errors and accumulated interchange errors have to be corrected by adjusting the controller settings according to procedures agreed upon by the whole interconnection. These accumulated errors as well as ACE serve as performance measures for LFC.

The main philosophy in the design of LFC is that each system should follow its own load very closely during normal operation, while during emergencies, each system should contribute according to its relative size in the interconnection without regard to the locality of the emergency. Thus, the most important factor in obtaining good control of a system is its inherent capability of following its own load. This is guaranteed if the system has adequate regulation margin as well as adequate response capability. Systems that have mainly thermal generation often have difficulty in keeping up with the load because of the slow response of the units.

The design of the controller itself is an important factor, and proper tuning of the controller parameters is needed to obtain "good" control without "excessive" movement of units. Tuning is system-specific, and although system simulations are often used as aids, most of the parameter adjustments are made in the field using heuristic procedures.

Economic Dispatch

Since all the generating units that are online have different costs of generation, it is necessary to find the generation levels of each of these units that would meet the load at the minimum cost. This has to take into account the fact that the cost of generation in one generator is not proportional to its generation level but is a nonlinear function of it. In addition, since the system is geographically spread out, the transmission losses are dependent on the generation pattern and must be considered in obtaining the optimum pattern.

Certain other factors have to be considered when obtaining the optimum generation pattern. One is that the generation pattern provide adequate reserve margins. This is often done by constraining the generation level to a lower boundary than the generating capability. A more difficult set of constraints to consider are the transmission limits. Under certain real-time conditions it is possible that the most economic pattern may not be feasible because of unacceptable line flows or voltage conditions. The present-day economic dispatch (ED) algorithm cannot handle these security constraints. However, alternative methods based on optimal power flows have been suggested but have not yet been used for real-time dispatch.

The minimum cost dispatch occurs when the incremental cost of all the generators is equal. The cost functions of the generators are nonlinear and discontinuous. For the equal marginal cost algorithm to work, it is necessary for them to be convex. These incremental cost curves are often represented as monotonically increasing piecewise-linear functions. A binary search for the optimal marginal cost is conducted by summing all the generation at a certain marginal cost and comparing it with the total power demand. If the demand is higher, a higher marginal cost is needed, and vice versa. This algorithm produces the ideal setpoints for all the generators for that particular demand, and this calculation is done every few minutes as the demand changes.

The losses in the power system are a function of the generation pattern, and they are taken into account by multiplying the generator incremental costs by the appropriate penalty factors. The penalty factor for each generator is a reflection of the sensitivity of that generator to system losses, and these sensitivities can be obtained from the transmission loss factors.

This ED algorithm generally applies to only thermal generation units that have cost characteristics of the type discussed here. The hydro units have to be dispatched with different considerations. Although there is no cost for the water, the amount of water available is limited over a period, and the displacement of fossil fuel by this water determines its worth. Thus, if the water usage limitation over a period is known, say from a previously computed hydro optimization, the water worth can be used to dispatch the hydro units.

LFC and the ED functions both operate automatically in realtime but with vastly different time periods. Both adjust generation levels, but LFC does it every few seconds to follow the load variation, while ED does it every few minutes to assure minimal cost. Conflicting control action is avoided by coordinating the control errors. If the unit control errors from LFC and ED are in the same direction, there is no conflict. Otherwise, a logic is set to either follow load (permissive control) or follow economics (mandatory control).

Reserve Monitoring

Maintaining enough reserve capacity is required in case generation is lost. Explicit formulas are followed to determine the spinning (already synchronized) and ready (10 min) reserves required. The availability can be assured by the operator manually, or, as mentioned previously, the ED can also reduce the upper dispatchable limits of the generators to keep such generation available.

Interchange Transaction Scheduling

The contractual exchange of power between utilities has to be taken into account by the LFC and ED functions. This is done by calculating the net interchange (sum of all the buy and sale agreements) and adding this to the generation needed in both the LFC and ED. Since most interchanges begin and end on the hour, the net interchange is ramped from one level to the new over a 10- or 20-min period straddling the hour. The programs achieve this automatically from the list of scheduled transactions.

Load Management

SCADA, with its relatively expensive RTUs installed at distribution substations, can provide status and measurements for distribution feeders at the substation. Distribution automation equipment is now available to measure and control at locations dispersed along distribution circuits. This equipment can monitor sectionalizing devices (switches, interruptors, fuses), operate switches for circuit reconfiguration, control voltage, read customers' meters, implement time-dependent pricing (on-peak, off-peak rates), and switch customer equipment to manage load. This equipment requires significantly increased functionality at distribution control centers.

Distribution control center functionality varies widely from company to company, and the following list is evolving rapidly.

- Data acquisition: Acquires data and gives the operator control over specific devices in the field. Includes data processing, quality checking, and storage.
- Feeder switch control: Provides remote control of feeder switches.
- Tagging and alarms: Provides features similar to SCADA.
- Diagrams and maps: Retrieves and displays distribution maps and drawings. Supports device selection from these displays. Overlays telemetered and operator-entered data on displays.
- Preparation of switching orders: Provides templates and information to facilitate preparation of instructions necessary to disconnect, isolate, reconnect, and reenergize equipment.
- Switching instructions: Guides operator through execution of previously prepared switching orders.
- Trouble analysis: Correlates data sources to assess scope of trouble reports and possible dispatch of work crews.
- Fault location: Analyzes available information to determine scope and location of fault.
- Service restoration: Determines the combination of remote control actions that will maximize restoration of service. Assists operator to dispatch work crews.
- Circuit continuity analysis: Analyzes circuit topology and device status to show electrically connected circuit segments (either energized or deenergized).
- Power factor and voltage control: Combines substation and feeder data with predetermined operating parameters to control distribution circuit power factor and voltage levels.
- Electrical circuit analysis: Performs circuit analysis, single-phase or three-phase, balanced or unbalanced.
- Load management: Controls customer loads directly through appliance switching (e.g., water heaters) and indirectly through voltage control.
- Meter reading: Reads customers' meters for billing, peak demand studies, time of use tariffs. Provides remote connect/disconnect.

Energy Management

Generation control and ED minimize the current cost of energy production and transmission within the range of available controls. Energy management is a supervisory layer responsible for economically scheduling production and transmission on a global basis and over time intervals consistent with cost optimization. For example, water stored in reservoirs of hydro plants is a resource that may be more valuable in the future and should, therefore, not be used now even though the cost of hydro energy is currently lower than thermal generation. The global consideration arises from the ability to buy and sell energy through the interconnected power system; it may be more economical to buy than to produce from plants under direct control. Energy accounting processes transaction information and energy measurements recorded during actual operation as the basis of payment for energy sales and purchases.

Energy management includes the following functions:

- System load forecast: Forecasts system energy demand each hour for a specified forecast period of 1 to 7 days.
- Unit commitment: Determines start-up and shut-down times for most economical operation of thermal generating units for each hour of a specified period of 1 to 7 days.
- Fuel scheduling: Determines the most economical choice of fuel consistent with plant requirements, fuel purchase contracts, and stockpiled fuel.
- Hydro-thermal scheduling: Determines the optimum schedule of thermal and hydro energy production for each hour of a study period up to 7 days while ensuring that hydro and thermal constraints are not violated.
- Transaction evaluation: Determines the optimal incremental and production costs for exchange (purchase and sale) of additional blocks of energy with neighboring companies.
- Transmission loss minimization: Recommends controller actions to be taken in order to minimize overall power system network losses.
- Security constrained dispatch: Determines optimal outputs of generating units to minimize production cost while ensuring that a network security constraint is not violated.
- Production cost calculation: Calculates actual and economical production costs for each generating unit on an hourly basis.

Security Control

Power systems are designed to survive all probable contingencies. A contingency is defined as an event that causes one or more important components such as transmission lines, generators, and transformers to be unexpectedly removed from service. Survival means the system stabilizes and continues to operate at acceptable voltage and frequency levels without loss of load. Operations must deal with a vast number of possible conditions experienced by the system, many of which are not anticipated in planning. Instead of dealing with the impossible task of analyzing all possible system states, security control starts with a specific state: the current state if executing the real-time network sequence; a postulated state if executing a study sequence. Sequence means sequential execution of programs that perform the following steps:

1. Determine the state of the system based on either current or postulated conditions.
2. Process a list of contingencies to determine the consequences of each contingency on the system in its specified state.
3. Determine preventive or corrective action for those contingencies which represent unacceptable risk.

Real-time and study network analysis sequences are diagramed in Fig. 12.3.

Security control requires topological processing to build network models and uses large-scale AC network analysis to determine system conditions. The required applications are grouped as a network subsystem that typically includes the following functions:

- Topology processor: Processes real-time status measurements to determine an electrical connectivity (bus) model of the power system network.
- State estimator: Uses real-time status and analog measurements to determine the “best” estimate of the state of the power system. It uses a redundant set of measurements; calculates voltages, phase angles, and power flows for all components in the system; and reports overload conditions.
- Power flow: Determines the steady-state conditions of the power system network for a specified generation and load pattern. Calculates voltages, phase angles, and flows across the entire system.
- Contingency analysis: Assesses the impact of a set of contingencies on the state of the power system and identifies potentially harmful contingencies that cause operating limit violations.

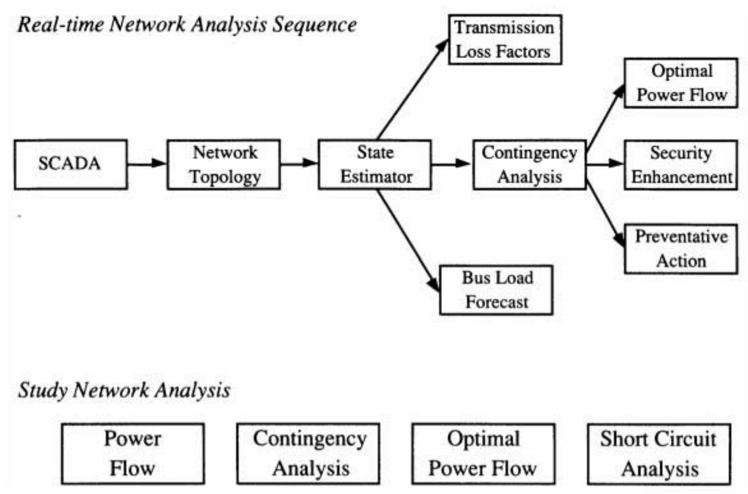


FIGURE 12.3 Real-time and study network analysis sequences.

- Optimal power flow: Recommends controller actions to optimize a specified objective function (such as system operating cost or losses) subject to a set of power system operating constraints.
- Security enhancement: Recommends corrective control actions to be taken to alleviate an existing or potential overload in the system while ensuring minimal operational cost.
- Preventive action: Recommends control actions to be taken in a "preventive" mode before a contingency occurs to preclude an overload situation if the contingency were to occur.
- Bus load forecasting: Uses real-time measurements to adaptively forecast loads for the electrical connectivity (bus) model of the power system network.
- Transmission loss factors: Determines incremental loss sensitivities for generating units; calculates the impact on losses if the output of a unit were to be increased by 1 MW.
- Short-circuit analysis: Determines fault currents for single-phase and three-phase faults for fault locations across the entire power system network.

Operator Training Simulator

Training simulators were originally created as generic systems for introducing operators to the electrical and dynamic behavior of power systems. Today, they model actual power systems with reasonable fidelity and are integrated with EMS to provide a realistic environment for operators and dispatchers to practice normal, every-day operating tasks and procedures as well as experience emergency operating situations. The various training activities can be safely and conveniently practiced with the simulator responding in a manner similar to the actual power system.

An operator training simulator (OTS) can be used in an investigatory manner to recreate past actual operational scenarios and to formulate system restoration procedures. Scenarios can be created, saved, and reused. The OTS can be used to evaluate the functionality and performance of new real-time EMS functions and also for tuning AGC in an off-line, secure environment.

The OTS has three main subsystems (Fig. 12.4).

Energy Control System

The energy control system (ECS) emulates normal EMS functions and is the only part of the OTS with which the trainee interacts. It consists of the supervisory control and data acquisition (SCADA) system, generation control system, and all other EMS functions.

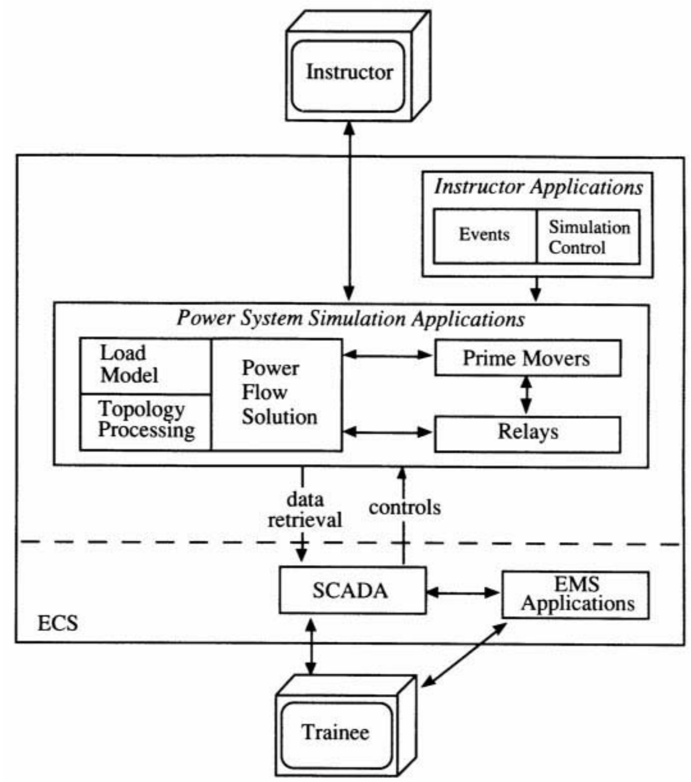


FIGURE 12.4 OTS block diagram.

Power System Dynamic Simulation

This subsystem simulates the dynamic behavior of the power system. System frequency is simulated using the "long-term dynamics" system model, where frequency of all units is assumed to be the same. The prime-mover dynamics are represented by models of the units, turbines, governors, boilers, and boiler auxiliaries. The network flows and states (bus voltages and angles, topology, transformer taps, etc.) are calculated at periodic intervals. Relays are modeled, and they emulate the behavior of the actual devices in the field.

Instructional System

This subsystem includes the capabilities to start, stop, restart, and control the simulation. It also includes making savecases, retrieving savecases, reinitializing to a new time, and initializing to a specific real-time situation.

It is also used to define event schedules. Events are associated with both the power system simulation and the ECS functions. Events may be deterministic (occur at a predefined time), conditional (based on a predefined set of power system conditions being met), or probabilistic (occur at random).

References

Application of Optimization Methods for Economy/Security Functions in Power System Operations, IEEE tutorial course, IEEE Publication 90EH0328-5-PWR, 1990.

Distribution Automation, IEEE Power Engineering Society, IEEE Publication EH0280-8-PBM, 1988.

C. J. Erickson, *Handbook of Electrical Heating,* IEEE Press, 1995.

Energy Control Center Design, IEEE tutorial course, IEEE Publication 77 TU0010-9 PWR, 1977.

Fundamentals of Load Management, IEEE Power Engineering Society, IEEE Publication EH0289-9-PBM, 1988.
Fundamentals of Supervisory Controls, IEEE tutorial course, IEEE Publication 91 EH0337-6 PWR, 1991.
M. Kleinpeter, *Energy Planning and Policy,* New York: Wiley, 1995.
Special issue on computers in power system operations, *Proc. IEEE,* 75, 12, 1987.
W. C. Turner, *Energy Management Handbook,* Fairmont Press, 1997.

Further Information

Current innovations and applications of new technologies and algorithms are presented in the following publications:

- *IEEE Power Engineering Review* (monthly)
- *IEEE Transactions on Power Systems* (bimonthly)
- *Proceedings of the Power Industry Computer Application Conference* (biannual)

12.2 Generation Control: Economic Dispatch and Unit Commitment

Charles W. Richter, Jr.

An area of power system control having a large impact on cost and profit is the optimal scheduling of generating units. A good schedule identifies which units to operate, and the amount to generate at each online unit in order to achieve a set of economic goals. These are the problems commonly referred to as the unit commitment (UC) problem, and the economic dispatch calculation, respectively. The goal is to choose a control strategy that minimizes losses (or maximizes profits), subject to meeting a certain demand and other system constraints. The following sections define EDC, the UC problem, and discuss methods that have been used to solve these problems. Realizing that electric power grids are complex interconnected systems that must be carefully controlled if they are to remain stable and secure, it should be mentioned that the tools described in this chapter are intended for steady-state operation. Short-term (less than a few seconds) changes to the system are handled by dynamic and transient system controls, which maintain secure and stable operation, and are beyond the scope of this discussion.

Economic Dispatch

Economic Dispatch Defined

An *economic dispatch calculation* (EDC) is performed to *dispatch*, or schedule, a set of online generating units to collectively produce electricity at a level that satisfies a specified demand in an economical manner. Each online generating unit may have many characteristics that make it unique, and which must be considered in the calculation. The amount of electricity demanded can vary quickly and the schedule produced by an EDC should leave units able to respond and adapt without major implications to cost or profit. The electric system may have limits (e.g., voltage, transmission, etc.) that impact the EDC and hence should be considered. Generating units may have prohibited generation levels at which resonant frequencies may cause damage or other problems to the system. The impact of transmission losses, congestion, and limits that may inhibit the ability to serve the load in a particular region from a particular generator (e.g., a low-cost generator) should be considered. The market structure within an operating region and its associated regulations must be considered in determining the specified demand, and in determining what constitutes economical operation. An independent system operator (ISO) tasked with maximizing social welfare would likely have a different definition of “economical” than does a generation company (GENCO) wishing to maximize its profit in a competitive environment. The EDC must consider all of these factors and develop a schedule that sets the generation levels in accordance with an economic objective function.

Factors to Consider in the EDC

The Cost of Generation

Cost is one of the primary characteristics of a generating unit that must be considered when dispatching units economically. The EDC is concerned with the short-term operating cost, which is primarily determined by fuel cost and usage. Fuel usage is closely related to generation level. Very often, the relationship between power level and fuel cost is approximated by a quadratic curve: $\mathbf{F} = \mathbf{aP}^2 + \mathbf{bP} + \mathbf{c}$. $\mathbf{c}$ is a constant term that represents the cost of operating the plant, $\mathbf{b}$ is a linear term that varies directly with the level of generation, and $\mathbf{a}$ is the term that accounts for efficiency changes over the range of the plant output. A quadratic relationship is often used in the research literature. However, due to varying conditions at certain levels of production (e.g., the opening or closing of large valves may affect the generation cost [Walters and Sheblé, 1992]), the actual relationship between power level and fuel cost may be more complex than a quadratic equation. Many of the long-term generating unit costs (e.g., costs attributed directly to starting and stopping the unit, capital costs associated with financing the construction) can be ignored for the EDC, since the decision to switch on, or *commit*, the units has already been made. Other characteristics of generating units that affect the EDC are the minimum and maximum generation levels at which they may operate. When binding, these constraints will directly impact the EDC schedule.

The Price

The price at which an electric supplier will be compensated is another important factor in determining an optimal economic dispatch. In many areas of the world, electric power systems have been, or still are, treated as a natural monopoly. Regulations allow the utilities to charge rates that guarantee them a nominal profit. In competitive markets, which come in a variety of flavors, price is determined through the forces of supply and demand. Economic theory and common sense tell us that if the total supply is high and the demand is low, the price is likely to be low, and vice versa. If the price is consistently below a GENCO's average total costs, the company may soon be bankrupt.

The Quantity Supplied

The amount of electric energy to be supplied is another fundamental input for the EDC. Regions of the world having regulations that limit competition often require electric utilities to serve all electric demand within a designated service territory. If a consumer switches on a motor, the electric supplier must provide the electric energy needed to operate the motor. In competitive markets, this *obligation to serve* is limited to those with whom the GENCO has a contract. Beyond its contractual obligations, the GENCO may be willing (if the opportunity arises) to supply additional consumer demand. Since the consumers have a choice of electric supplier, a GENCO determining the schedule of its own online generating units may choose to supply all, none, or only a portion of that additional consumer demand. The decision is dependent on the objective of the entity performing the EDC (e.g., profit maximization, improving reliability, etc.).

EDC and System Limitations

A complex network of transmission and distribution lines and equipment are required to move the electric energy from the generating units to the consumer loads. The secure operation of this network depends on bus voltage magnitudes and angles being within certain tolerances. Excessive transmission line loading can also affect the security of the power system network. Since superconductivity is a relatively new field, lossless transmission lines are expensive and are not commonly used. Therefore, some of the energy being transmitted over the system is converted into heat and is consequently lost. The schedule produced by the EDC directly affects losses and security; hence, constraints ensuring proper system operation must be considered when solving the EDC problem.

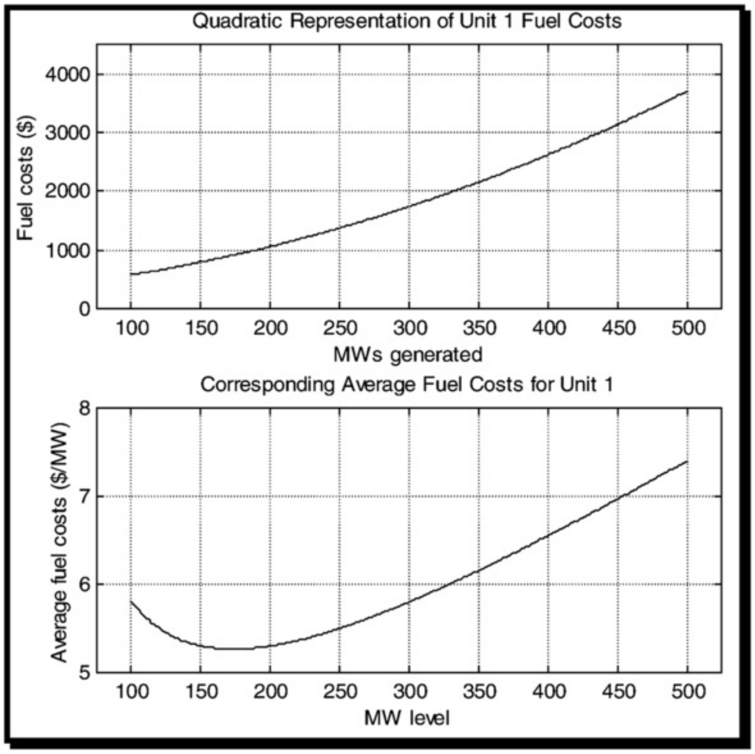


FIGURE 12.5 Relationship between fuel input and power output.

The Objective of EDC

In a regulated, vertically integrated, monopolistic environment, the obligated-to-serve electric utility performs the EDC for the entire service area by itself. In such an environment, providing electricity in an "economical manner" means minimizing the cost of generating electricity, subject to meeting all demand and other system operating constraints. In a competitive environment, the way an EDC is done can vary from one market structure to another. For instance, in a decentralized market, the EDC may be performed by a single GENCO wishing to maximize its expected profit given the prices, demands, costs, and other constraints described above. In a power pool, a central coordinating entity may perform an EDC to centrally dispatch generation for many GENCOs. Depending on the market rules, the generation owners may be able to mask the cost information of their generators. In this case, bids would be submitted for various price levels and used in the EDC.

The Traditional EDC Mathematical Formulation

Assuming operation under a vertically integrated, monopolistic environment, we must meet all demand, D. We must also consider minimum and maximum limits for each generating unit, $P_i^{\min}$ and $P_i^{\max}$. We will assume that the fuel costs of the ith operating plant may be modeled by a quadratic equation as shown in Eq. (12.1), and shown graphically in Fig. 12.5. Note that the average fuel costs are also shown in Fig. 12.5.

$$F_i = a_i P_i^2 + b_i P_i + c_i \qquad \left(\text{fuel costs of ith generator}\right) \tag{12.1}$$

Thus, for N online generating units, we can write a Lagrangian equation, L, which describes the total cost and associated demand constraint, D.

$$L = F_T + \lambda\left(D - \sum_{i=1}^{N} P_i\right) = \sum_{i=1}^{N}\left(a_i P_i^2 + b_i P_i + c_i\right) + \lambda\cdot\left(D - \sum_{i=1}^{N} P_i\right)$$

$$F_T = \sum_{i=1}^{N} F_i \qquad \text{(Total fuel cost is a summation of costs for all online plants)} \tag{12.2}$$

$$P_i^{\min} \le P_i \le P_i^{\max} \qquad \text{(Generation must be set between the min and max amounts)}$$

Additionally, note that c_i is a constant term that represents the cost of operating the ith plant, b_i is a linear term that varies directly with the level of generation, P_i, and a_i are terms that account for efficiency changes over the range of the plant output.

In this example, the objective will be to minimize the cost of supplying demand with the generating units that are online. From calculus, a minimum or a maximum can be found by taking the N + 1 derivatives of the Lagrangian with respect to its variables, and setting them equal to zero. The shape of the curves is often assumed well behaved — monotonically increasing and convex — so that determining the second derivative is unnecessary.

$$\frac{\partial L}{\partial P_i} = 2a_i P_i + b_i - \lambda = 0 \Rightarrow \lambda = 2a_i P_i + b_i \tag{12.3}$$

$$\frac{\partial L}{\partial \lambda} = \left(D - \sum_{1}^{N} P_i\right) = 0 \tag{12.4}$$

λ_i is the commonly used symbol for the "marginal cost" of the *i*-th unit. At the margin of operation, the marginal cost tells us how many additional dollars the GENCO will have to spend to increase the generation by an additional MW. The marginal cost curve is an positively sloped line if a quadratic equation is being used to represent the fuel curve of the unit. The higher the quantity being produced, the greater the cost of adding an additional unit of the goods being produced. Economic theory says that if a GENCO has a set of plants and it wants to increase production by one unit, it should increase production at the plant that provides the most benefit for the least cost. The GENCO should do this until that plant is no longer providing the greatest benefit for a given cost. At that point it finds the plant now giving the highest benefit-to-cost ratio and increases its production. This is done until all plants are operating at the same marginal cost. When all unconstrained online plants have the same marginal cost, λ (i.e., $\lambda_1 = \lambda_2 = \ldots = \lambda_i = \ldots = \lambda_{SYSTEM}$), then the cost is at a minimum for that amount of generation. If there were binding constraints, it would prevent the GENCO from achieving that scenario.

If a constraint is binding on a particular unit (e.g., P_i becomes $P_i^{\max}$ when attempting to increase production), the marginal cost of that unit is considered to be infinite. No matter how much money is available to increase plant production by one unit, it cannot do so. (Of course, in the long term, things may be done that can reduce the effect of the constraint, but that is beyond the scope of this discussion.)

EDC Solution Techniques

There are many ways to obtain the optimum power levels that will achieve the objective for the EDC problem being considered. For very simple situations, one may solve the solution directly; but when the number of constraints that introduce nonlinearities to the problem grows, iterative search techniques become necessary. Wood and Wollenberg (1996) describe many such methods of calculating economic dispatch, including the graphical technique, the lambda-iteration method, and the first- and second-order gradient methods. Another method that works well, even when fuel costs are not modeled by a simple quadratic equation, is the genetic algorithm.

TABLE 12.1 Generator Data and Solution for EDC Example

Unit Number	Unit Parameters					Solution		
	P_{min}	P_{max}	A	B	C	P_i (MW)	\$/MW (λ_i)	Cost \$/hour
1	100	500	.01	1.8	300	233.2456	6.4649	1263.90
2	50	300	.012	2.24	210	176.0380	6.4649	976.20
3	100	400	.006	2.35	290	342.9094	6.4649	1801.40
4	100	500	.008	2.5	340	247.8070	6.4649	1450.80

In highly competitive scenarios, each inaccuracy in the model can result in losses to the GENCO. A very detailed model might include many nonlinearities, (e.g., valve-point loading, prohibited regions of operation, etc.). Such nonlinearities may mean that it is not possible to calculate a derivative. If the relationship is not well-behaved, there may be no proof that the solution can ever be optimal. With greater detail in the model comes an increase in the amount of time to perform the EDC. Since the EDC is performed quite frequently (on the order of every few minutes), and because it is a real-time calculation, the solution technique should be quick. Since an inaccurate solution may produce a negative impact on the company profits, the solution should also be accurate.

An Example of Cost Minimizing EDC

To illustrate how the EDC is solved via the graphical method, an example is presented here. Assume that a GENCO needs to supply 1000 MW of consumer demand, and that Table 12.1 describes the system on-line units that it is dispatching in a traditional, i.e., vertically integrated, monopolistic environment. Figure 12.6 shows the marginal costs of each of the units over their entire range. It also shows an aggregated marginal cost curve that could be called the system marginal cost curve. This aggregated system curve was created by a horizontal summation of the four individual graphs. Once the system curve is created, one simply finds the desired power level (i.e., 1000 MW) along the x-axis. Follow it up to the curve, and then look to the left. On the y-axis, the system marginal cost can be read. Since no limits were reached, each of the individual λ_is is the same as the system λ. The GENCO can find the λ_i

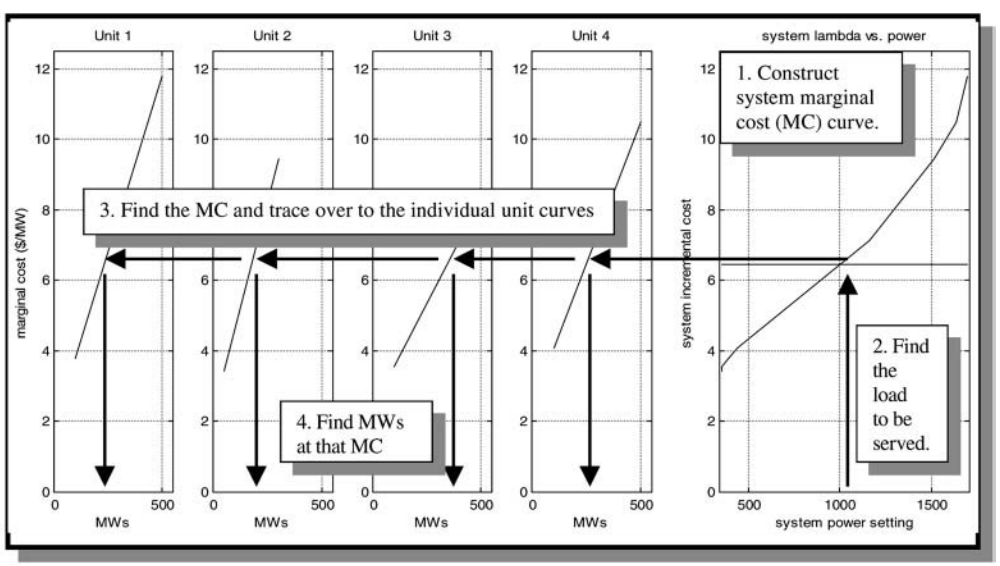


FIGURE 12.6 Unit and aggregated marginal cost curves for solving EDC with the graphical method.

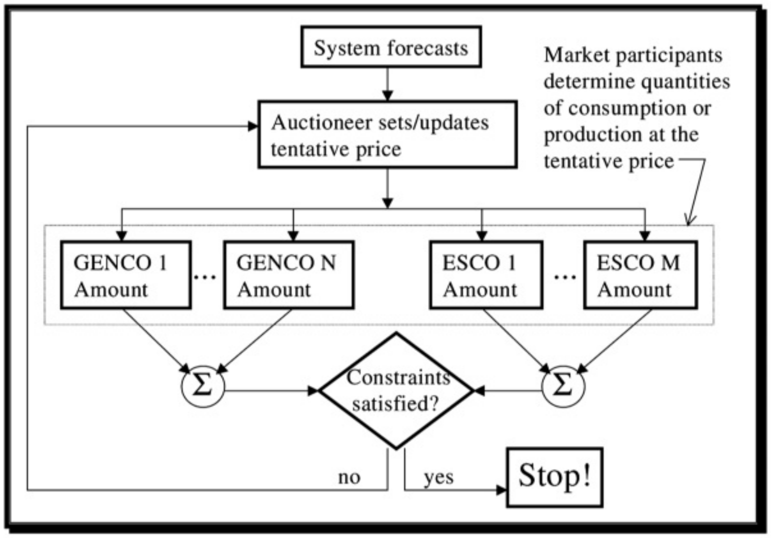


FIGURE 12.7 Economic dispatch and/or unit commitment as an auction.

on each of the unit curves and draw a line straight down from the point where the marginal cost, λ, crosses the curve to find its power level. The generation levels of each online unit are easily found and the solution is shown in the right-hand columns of Table 12.1. The procedure just described is the graphical method of EDC. If the system marginal cost had been above the diagonal portion of an individual unit curve, then we simply set that unit at its P^{max}.

EDC and Auctions

Competitive electricity markets vary in their operating rules, social objectives, and in the mechanism they use to allocate prices and quantities to the participants. Commonly, an auction is used to match buyers with sellers and to achieve a price that is considered fair. Auctions can be sealed bid, open outcry, ascending ask English auctions, descending ask Dutch auctions, etc. Regardless of the solution technique used to find the optimal allocation, the economic dispatch is essentially performing the same allocation that an auction would. Suppose an auctioneer were to call out a price, and ask the participating/online generators how much power they would generate at that level. The reply amounts could be summed to determine the production level at that price. If all of the constraints, including demand, are met, then the most economical dispatch has been achieved. If not, the auctioneer adjusts the price and asks for the amounts at the new price. This procedure is repeated until the constraints are satisfied. Prices may ascend as in the English auction, or they may descend as in the Dutch auction. See Fig. 12.7 for a graphical depiction of this process. For further discussion on this topic, the interested reader is referred to Sheblé (1999).

The Unit Commitment Problem

Unit Commitment Defined

The *unit commitment* (UC) problem is defined as the scheduling of a set of generating units to be on, off, or in stand-by/banking mode for a given period of time to meet a certain objective. For a power system operated by a vertically integrated monopoly, committing units is performed centrally by the utility, and the objective is to minimize costs subject to supplying all demand (and reserve margins). In a competitive environment, each GENCO must decide which units to commit, such that profit is maximized, based on the number of contracted MW; the additional MWhr it forecasts that it can profitably wrest from its competitors in the spot market; and the prices at which it will be compensated.

UC Schedule

Hour	1 2 3 4 5 6 ... T
Gen#1:	1 1 1 1 1 1 ... 0
Gen#2:	0 0 0 1 1 1 ... 1
Gen#3:	1 1 1 0 0 0 ... 1
....	
Gen#N:	1 1 1 1 1 1 ... 0

0=unit off-line 1=unit on-line

FIGURE 12.8 A typical unit commitment schedule.

A UC schedule is developed for N units and T periods. A typical UC schedule might look like the one shown in Fig. 12.8. Since uncertainty in the inputs becomes large beyond one week into the future, the UC schedule is typically developed for the following week. It is common to consider schedules that allow unit-status change from hour to hour, so that a weekly schedule is made up of 168 periods. In finding an optimal schedule, one must consider fuel costs, which can vary with time, start-up and shut-down costs, maximum ramp rates, the minimum up-times and minimum down-times, crew constraints, transmission limits, voltage constraints, etc. Because the problem is discrete, the GENCO may have many generating units, a large number of periods may be considered, and because there are many constraints, finding an optimal UC is a complex problem.

Factors to Consider in Solving the UC Problem

The Objective of Unit Commitment

The objective of the unit commitment algorithm is to schedule units in the most economical manner. For the GENCO deciding which units to commit in the competitive environment, economical manner means one that maximizes its profits. For the monopolist operating in a vertically integrated electric system, economical means minimizing the costs.

The Quantity to Supply

In systems with vertically integrated monopolies, it is common for electric utilities to have an obligation to serve all demand within their territory. Forecasters provide power system operators an estimated amount of power demanded. The UC objective is to minimize the total operational costs subject to meeting all of this demand (and other constraints they may be considering).

In competitive electric markets, the GENCO commits units to maximize its profit. It relies on spot and forward bilateral contracts to make part of the total demand known *a priori*. The remaining share of the demand that it may pick up in the spot market must be predicted. This market share may be difficult to predict since it depends on how its price compares to that of other suppliers.

The GENCO may decide to supply less demand than it is physically capable of. In the competitive environment, the obligation to serve is limited to those with whom the GENCO has a contract. The GENCO may consider a schedule that produces less than the forecasted demand. Rather than switching on an additional unit to produce one or two unsatisfied MW, it can allow its competitors to provide that 1 or 2 MW that might have substantially increased its average costs.

Compensating the Electricity Supplier

Maximizing profits in a competitive environment requires that the GENCO know what revenue is being generated by the sale of electricity. While a traditional utility might have been guaranteed a fixed rate of return based on cost, competitive electricity markets have varying pricing schemes that may price electricity at the level of the last accepted bid, the average of the buy, ask, and sell offer, etc. When submitting offers to an auctioneer, the GENCO's offer price should reflect its prediction market share,

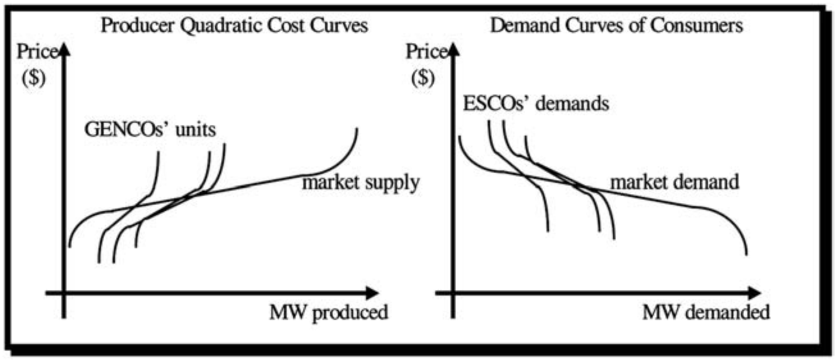


FIGURE 12.9 Treating the market as an additional generator and/or load.

since that determines how many units they have switched on, or in banking mode. GENCOs recovering costs via prices set during the bidding process will note that the UC schedule directly affects the average cost, which indirectly affects the offering price, making it an essential input to any successful bidding strategy.

Demand forecasts and expected market prices are important inputs to the profit-based UC algorithm; they are used to determine the expected revenue, which in turn affects the expected profit. If a GENCO produces two UC schedules each having different expected costs and different expected profits, it should implement the one that provides for the largest profit, which will not necessarily be the one that costs the least. Since prices and demand are so important in determining the optimal UC schedule, price prediction and demand forecasts become crucial. An easy-to-read description of the cost-minimizing UC problem and a stochastic solution that considers spot markets has been presented in Takriti, Krasenbrink, and Wu (1997).

The Source of Electric Energy

A GENCO may be in the business of electricity generation, but it should also consider purchasing electricity from the market, if it is less expensive than its own generating unit(s). The existence of liquid markets gives energy trading companies an additional source from which to supply power that may not be as prevalent in monopolistic systems. See Fig. 12.9. To the GENCO, the market supply curve can be thought of as a pseudo-unit to be dispatched. The supply curve for this pseudo-unit represents an aggregate supply of all of the units participating in the market at the time in question. The price forecast essentially sets the parameters of the unit. This pseudo-unit has no minimum uptime, minimum downtime, or ramp constraints; there are no direct start-up and shutdown costs associated with dispatching the unit.

The liquid markets that allow the GENCO to schedule an additional pseudo unit, also act as a load to be supplied. The total energy supplied should consist of previously arranged bilateral or multilateral contracts arranged through the markets (and their associated reserves and losses). While the GENCO is determining the optimal unit commitment schedule, the energy demanded by the market (i.e., market demand) can be represented as another DISTCO or ESCO buying electricity. Each entity buying electricity should have its own demand curve. The market demand curve should reflect the aggregate of the demand of all the buying agents participating in the market.

Mathematical Formulation for UC

The mathematical formulation for UC depends upon the objective and the constraints that are considered important. Traditionally, the monopolist cost-minimization UC problem has been formulated (Sheblé, 1985):

$$\text{Minimize F} = \sum_{n}^{N}\sum_{t}^{T}\left[\left(C_{nt} + MAINT_{nt}\right)\cdot U_{nt} + SUP_{nt}\cdot U_{nt}\left(1-U_{nt}\right) + SDOWN_{nt}\cdot\left(1-U_{nt}\right)\cdot U_{nt-1}\right] \quad (12.5)$$

subject to the following constraints:

$$\sum_{n}^{N}\left(U_{nt}\cdot P_{nt}\right)=D_t \qquad \text{(demand constraint)}$$

$$\sum_{n}^{N}\left(U_{nt}\cdot P\max_n\right)\geq D_t+R_i \qquad \text{(capacity constraint)}$$

$$\sum_{n}^{N}\left(U_{nt}\cdot Rs\max_n\right)\geq R_t \qquad \text{(system reserve constraint)}$$

When formulating the profit-maximizing UC problem for a competitive environment, the obligation-to-serve is gone. The demand constraint changes from an equality to an inequality (≤). In the formulation presented here, we lump the reserves in with the demand. Essentially we are assuming that buyers are required to purchase a certain amount of reserves per contract. In addition to the above changes, formulating the UC problem for the competitive GENCO changes the objective function from cost minimization to profit maximization as shown in Eq. (12.6) below. The UC solution process is shown in block diagram form in Fig. 12.10.

$$\text{Max }\Pi=\sum_{n}^{N}\sum_{t}^{T}\left(\mathrm{P}_{\mathrm{nt}}\cdot fp_t\right)\cdot U_{nt}-F \tag{12.6}$$

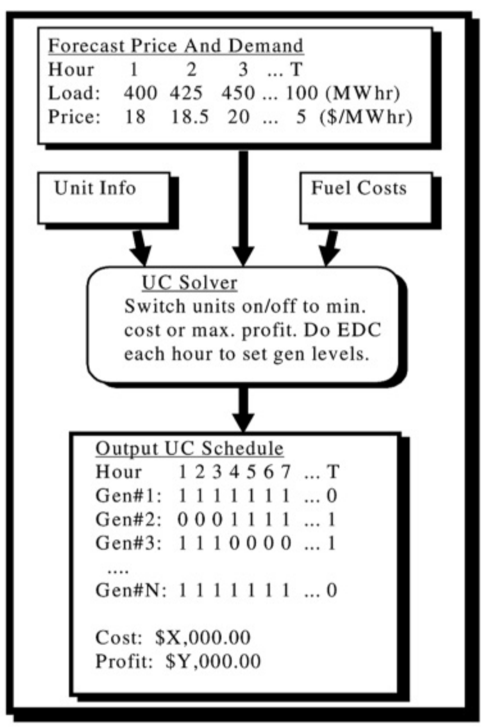


FIGURE 12.10 Block diagram of the UC solution process.

subject to:

$$D_t^{contracted} \le \sum_{n}^{N} \left(U_{nt} \cdot P_{nt} \right) \le D_t' \qquad \left(\text{demand constraint w/out obligation-to-serve}\right)$$

$$P\min_n \le P_{nt} \le P\max_n \qquad \left(\text{capacity limits}\right)$$

$$\left| P_{nt} - P_{n,\,t\text{-}1} \right| \le \text{Ramp}_n \qquad \left(\text{ramp rate limits}\right)$$

where individual terms are defined as follows:

U_{nt} = up/down time status of unit n at time period t (U_{nt} = 1 unit on, U_{nt} = 0 unit off)
P_{nt} = power generation of unit n during time period t
D_t = load level in time period t
D_t' = forecasted demand at period t (includes reserves)
$D_t^{contract}$ = contracted demand at period t (includes reserves)
fp_t = forecasted price/MWhr for period t
R_t = system reserve requirements in time period t
C_{nt} = production cost of unit n in time period t
SUP_{nt} = start-up cost for unit n, time period t
$SDOWN_{nt}$ = shut-down cost for unit n, time period t
$MAINT_{nt}$ = maintenance cost for unit n, time period t
N = number of units
T = number of time periods
$Pmin_n$ = generation low limit of unit n
$Pmax_n$ = generation high limit of unit n
$Rsmax_n$ = maximum contribution to reserve for unit n

Although it may happen in certain cases, the schedule that minimizes cost is not necessarily the schedule that maximizes profit. Providing further distinction between the cost-minimizing UC for the monopolist and the profit maximizing competitive GENCO is the obligation-to-serve; the competitive GENCO may choose to generate less than the total consumer demand. This allows a little more flexibility in the UC schedules. In addition, our formulation assumes that prices fluctuate according to supply and demand. In cost-minimizing paradigms, it is assumed that leveling the load curve helps to minimize the cost. When maximizing profit, the GENCO may find that under certain conditions, it may profit more under a non-level load curve. The profit depends not only on cost, but also on revenue. If revenue increases more than the cost does, the profit will increase.

The Importance of EDC to the UC Solution

The economic dispatch calculation (EDC) is an important part of UC. It is used to assure that sufficient electricity will be available to meet the objective each hour of the UC schedule. For the monopolist in a vertically integrated environment, EDC will set generation so that costs are minimized subject to meeting the demand. For the price-based UC, the price-based EDC adjusts the power level of each online unit until each has the same incremental cost (i.e., $\lambda_1 = \lambda_2 = \ldots = \lambda_i = \ldots = \lambda_T$). If a GENCO is operating in a competitive framework that requires its bids to cover fixed, start-up, shutdown, and other costs associated with transitioning from one state to another, then the incremental cost used by EDC must embed these costs. We shall refer to this modified marginal cost as a pseudo λ. The competitive generator will generate if the pseudo λ is less than or equal to the competitive price. A simple way to allocate the fixed and transitional costs that result in a \$/MWhr figure is shown in Eq. (12.7):

$$\lambda_t = fp_t - \frac{\sum_t \sum_n (\textit{transition costs}) + \sum_t \sum_n (\textit{fixed costs})}{\sum_t^T \sum_n^N P_{nt}} \tag{12.7}$$

Other allocation schemes that adjust the marginal cost/price according to the time of day or price of power would be just as easy to implement and should be considered in building bidding strategies. Transition costs include start-up, shutdown, and banking costs, and fixed costs (present for each hour that the unit is on), which would be represented by the constant term in the typical quadratic cost curve approximation. For the results presented later in this chapter, we approximate the summation of the power generated by the forecasted demand.

The competitive price is assumed to be equal to the forecasted price. If the GENCO's supply curve is indicative of the system supply curve, then the competitive price will correspond to the point where the demand and supply curves cross. EDC sets the generation level corresponding to the point where the GENCO's supply curve crosses the demand curve, or to the point where the forecasted price is equal to the supply curve, whichever is lower.

Solution Methods

Solving the UC problem to find an optimal solution can be difficult. The problem has a large solution space that is discrete and nonlinear. As mentioned above, solving the UC problem requires that many economic dispatch calculations be performed. One possible way to determine the optimal schedule is to do an exhaustive search. Exhaustively considering all possible ways that units can be switched on or off for a small system can be done, but for a reasonably sized system this would take too long. Solving the UC problem for a realistic system generally involves using methods like Lagrangian relaxation, dynamic programming, genetic algorithms, or other heuristic search techniques. The interested reader may find many useful references regarding cost-minimizing UC for the monopolist in Sheblé and Fahd (1994) and Wood and Wollenberg (1996). Another heuristic technique that has shown much promise and that offers many advantages (e.g., time-to-solution for large systems and ability to simultaneously generate multiple solutions) is the genetic algorithm.

A Genetic-Based UC Algorithm

The Basics of Genetic Algorithms

A genetic algorithm (GA) is a search algorithm often used in nonlinear discrete optimization problems. The development of GAs was inspired by the biological notion of evolution. Initially described by John Holland, they were popularized by David Goldberg who described the basic genetic algorithm very well (Goldberg, 1989). In a GA, data, initialized randomly in a data structure appropriate for the solution to the problem, evolves over time and becomes a suitable answer to the problem. An entire population of candidate solutions (data structures with a form suitable for solving for the problem being studied) is "randomly" initialized and evolves according to GA rules. The data structures often consist of strings of binary numbers that are mapped onto the solution space for evaluation. Each solution (often termed a creature) is assigned a fitness — a heuristic measure of its quality. During the evolutionary process, those creatures having higher fitness are favored in the parent selection process and are allowed to procreate. The parent selection is essentially a random selection with a fitness bias. The type of fitness bias is determined by the parent selection method. Following the parent selection process, the processes of crossover and mutation are utilized and new creatures are developed that ideally explore a different area of the solution space. These new creatures replace less fit creatures from the existing population. Figure 12.11 shows a block diagram of the general GA.

GA for Price-Based UC

The algorithm presented here solves the UC problem for the profit maximizing GENCO operating in the competitive environment (Richter et al., 1999). Research reveals that various GAs have been used by

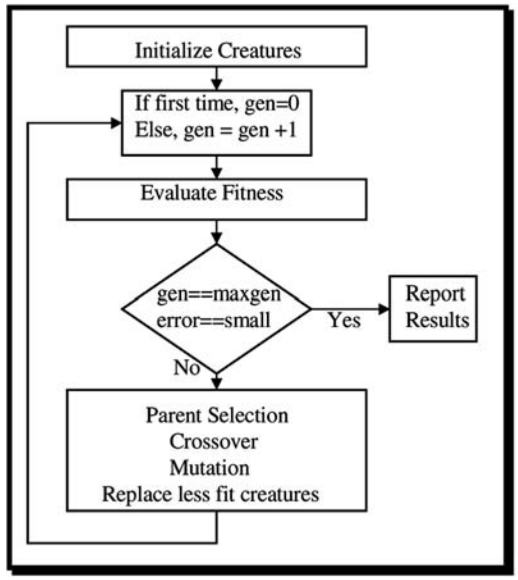


FIGURE 12.11 A simple genetic algorithm.

many researchers in solving the UC problem (Kondragunta, 1997; Kazarlis et al., 1995). However, the algorithm presented here is a modification of a genetic-based UC algorithm for the cost-minimizing monopolist described in Maifeld and Sheblé (1996). Most of the modifications are to the fitness function, which no longer rewards schedules that minimize cost, but rather those that maximize profit. The intelligent mutation operators are preserved in their original form. The schedule format is the same. The algorithm is shown in block diagram format in Fig. 12.12.

The algorithm first reads in the contract demand and prices, the forecast of remaining demand, and forecasted spot prices (which are calculated for each hour by another routine not described here). During the initialization step, a population of UC schedules is randomly initialized. See Fig. 12.13. For each member of the population, EDC is called to set the level of generation of each unit. The cost of each schedule is calculated from the generator and data read in at the beginning of the program. Next, the fitness (i.e., the profit) of each schedule in the population is calculated. "Done?" checks to see whether the algorithm as either cycled through for the maximum number of generations allowed, or whether other stopping criteria have been met. If done, then the results are written to a file; if not done, the algorithm goes to the reproduction process.

During reproduction, new schedules are created. The first step of reproduction is to select parents from the population. After selecting parents, candidate children are created using two-point crossover as shown in Fig. 12.14. Following crossover, standard mutation is applied. Standard mutation involves turning a randomly selected unit on or off within a given schedule.

An important feature of the previously developed UC-GA (Maifeld and Sheblé, 1996) is that it spends as little time as possible doing EDC. After standard mutation, EDC is called to update the profit only for the mutated hour(s). An hourly profit number is maintained and stored during the reproduction process, which dramatically reduces the amount of time required to calculate the profit over what it would be if EDC had to work from scratch at each fitness evaluation. In addition to the standard mutation, the algorithm uses two "intelligent" mutation operators that work by recognizing that, because of transition costs and minimum uptime and downtime constraints, 101 or 010 combinations are undesirable. The first of these operators would purge this undesirable combination by randomly changing 1s to 0s or vice versa. The second of these intelligent mutation operators purges the undesirable combination by changing 1 to 0 or 0 to 1 based on which of these is more helpful to the profit objective.

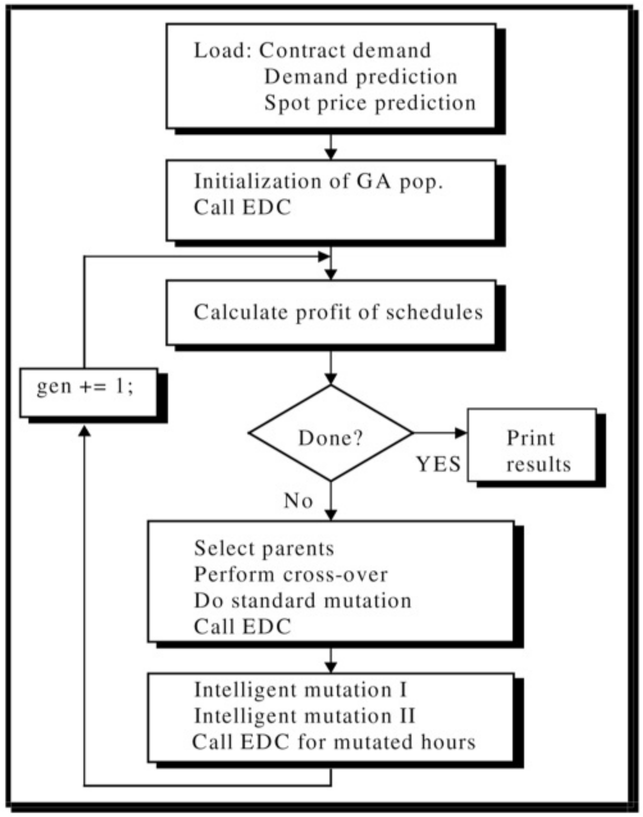


FIGURE 12.12 GA-UC block diagram.

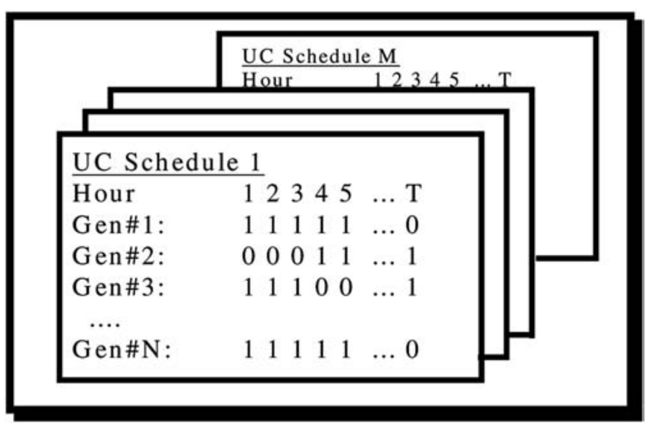


FIGURE 12.13 A population of UC schedules.

Price-Based UC-GA results

The UC-GA is run on a small system so that its solution can be easily compared to a solution by exhaustive search. Before running the UC-GA, the GENCO needs to first get an accurate hourly demand and price forecast for the period in question. Developing the forecasted data is an important topic, but beyond the scope of our analysis. For the results presented in this section, the forecasted load and prices are taken to be those shown in Table 12.2. In addition to loading the forecasted hourly price and demand, the

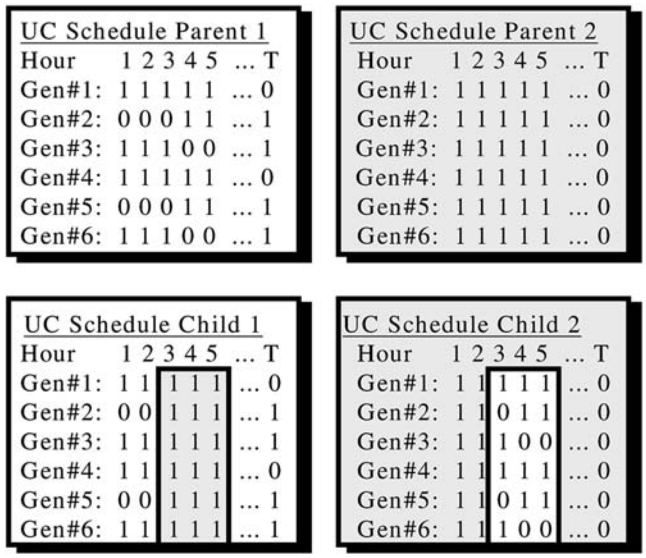


FIGURE 12.14 Two-point crossover on UC schedules.

TABLE 12.2 Forecasted Demand and Prices for 2-Generator Case

Hour	Load Forecast (MWhr)	Price Forecast ($/MWhr)	Hour	Load Forecast (MWhr)	Price Forecast ($/MWhr)
1	285	25.87	8	328	8.88
2	293	23.06	9	326	9.12
3	267	19.47	10	298	8.88
4	247	18.66	11	267	25.23
5	295	21.38	12	293	26.45
6	292	12.46	13	350	25.00
7	299	9.12	14	350	24.00

TABLE 12.3 Unit Data for 2-Generator Case

	Generator 0	Generator 1
Pmin (MW)	40	40
Pmax (MW)	180	180
A (constant)	58.25	138.51
B (linear)	8.287	7.955
C (quadratic)	7.62e-06	3.05e-05
Bank cost ($)	192	223
Start-up cost($)	443	441
Shut-down cost($)	750	750
Min-uptime (hr)	4	4
Min-downtime (hr)	4	4

UC-GA program needs to load the parameters of each generator to be considered. We are modeling the generators with a quadratic cost curve (e.g., $A + B(P) + C(P)^2$), where P is the power level of the unit. The data for the 2-generator case is shown in Table 12.3.

In addition to the 2-unit cases, a 10-unit, 48-hour case is included in this chapter to show that the GA works well on larger problems. While dynamic programming quickly becomes too computationally expensive to solve, the GA scales up linearly with number of hours and units. Figure 12.15 shows the

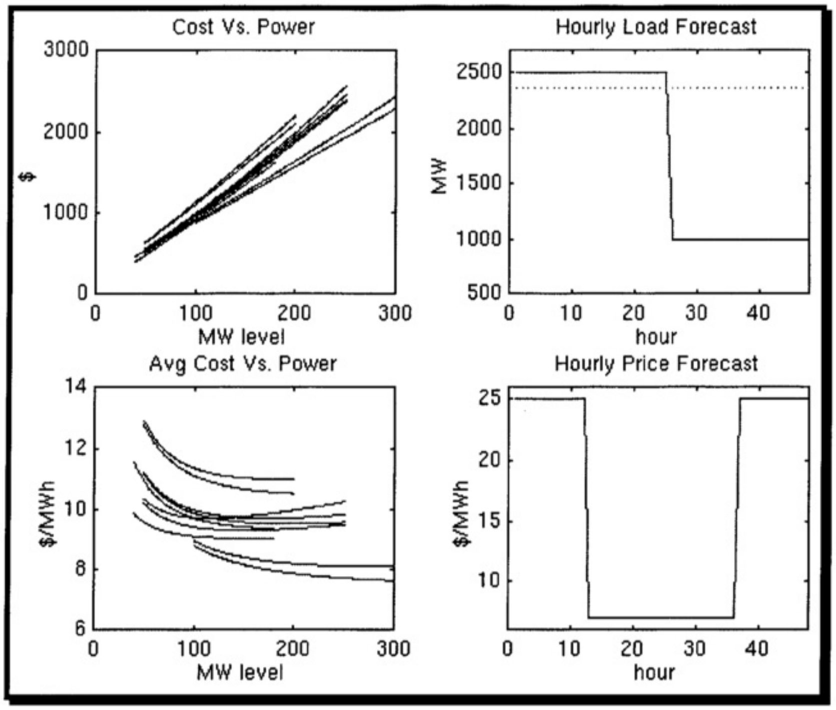


FIGURE 12.15 Data for 10-unit, 48-hour case.

TABLE 12.4 GA Control Parameters

Parameter	Setpoint	Parameter	Setpoint
# of Units	2	System reserve (%)	10
# of Hours	10	Children per generation	10
Popsize	20	UC schedules to keep	1
Generations	50	Random number seed	0.20

costs and average costs (without transition costs) of the 10 generators, as well as the hourly price and load forecasts for the 48 hours. The data was chosen so that the optimal solution was known *a priori*. The dashed line in the load forecast represents the maximum output of the 10 units.

Before running the UC-GA, the user specifies the control parameters shown in Table 12.4, including the number of generating units and number of hours to be considered in the study. The "popsize" is the size of the GA population. The execution time varies approximately linearly with the popsize. The number of generations indicates how many times the GA will go through the reproduction phase. System reserve is the percentage of reserves that the buyer must maintain for each contract. Children per generation tells us how much of the population will be replaced each generation. Changing this can affect the convergence rate. If there are multiple optima, faster convergence can trap the GA in a local suboptimal solution. "UC schedules to keep" indicates the number of schedules to write to file when finished. There is also a random number seed that is set between 0 and 1.

In the 2-generator test cases, the UC-GA was run for the units listed in Table 12.3, and for the forecasted loads and prices listed in Table 12.2. The parameters listed in Table 12.4 were adjusted accordingly. To ensure that the UC-GA is finding optimal solutions, an exhaustive search was performed on some of the smaller cases. Table 12.5 shows the time to solution in seconds for the UC-GA and the exhaustive search methods. For small cases, the exhaustive search was performed and solution time compared to that of

TABLE 12.5 Comparing UC-GA with Exhaustive Search

No. of Generators in Schedule	No. of Hours in Schedule	GA Finds Optimal Solution?	Solution Time for GA (s)	Solution Time Exhaustive Search (s)
2	10	Yes	0.5	674
2	12	Yes	2	6482
2	14	Yes	10	(estimated) 62340
10	48	Yes	730	(estimated) 2E138

TABLE 12.6 The Best UC-GA Schedules of the Population

	Best Schedule for 2-Unit, 10-Hour Case
Unit 1	1111100000
Unit 2	0000000000
Cost	$17,068.20
Profit	$2,451.01
	Best Schedule for 2-Unit, 12-Hour Case
Unit 1	111111000011
Unit 2	000000000000
Cost	$24,408.50
Profit	$4,911.50
	Best Schedule Found by UC-GA for 10-Unit, 48-Hour Case
Unit 1	111111111111000000000000000000000000111111111111
Unit 2	111111111111000000000000000000000000000000000000
Unit 3	111111111111000000000000000000000000000000000000
Unit 4	111111111111000000000000000000000000000000000000
Unit 5	111111111111000000000000000000000000000000000000
Unit 6	111111111111000000000000000000000000000000000000
Unit 7	111111111111000000000000000000000000111111111111
Unit 8	111111111111000000000000000000000000000000000000
Unit 9	111111111111000000000000000000000000111111111111
Unit 10	111111111111000000000000000000000000111111111111
Cost	$325,733.00
Profit	$676,267.00

the UC-GA. Since the exhaustive search solution times were estimated to be prohibitively lengthy, the latter cases were not compared against exhaustive search solutions. Cases with known optimal solutions were used to verify that the UC-GA was, in fact, working for the large cases.

Table 12.6 shows the optimal UC schedules found by the UC-GA for selected cases. Figure 12.16 shows the maximum, minimum and average fitnesses (profit) during each generation of the UC-GA on the 2-generator, 14-hour/period case. The best individual of the population climbs quite rapidly to near the optimal solution. Half of the population is replaced each generation; often the child solutions are poor solutions, hence the minimum fitness tends to remain low over the generations, which is typical for GA optimization.

In the schedules shown in Table 12.6, it may appear as though minimum up- and downtime constraints are being violated. When calculating the cost of such a schedule, the algorithm ensures that the profit is based on a valid schedule by considering a zero surrounded by ones to be a banked unit, and so forth. In addition, note that only the best solution of the population for each of the cases is shown. The existence of additional valid solutions, which may have been only slightly suboptimal in terms of profit, is one of the main advantages of using the GA. It gives the system operator the flexibility to choose the best schedule from a group of schedules to accommodate things like forced maintenance.

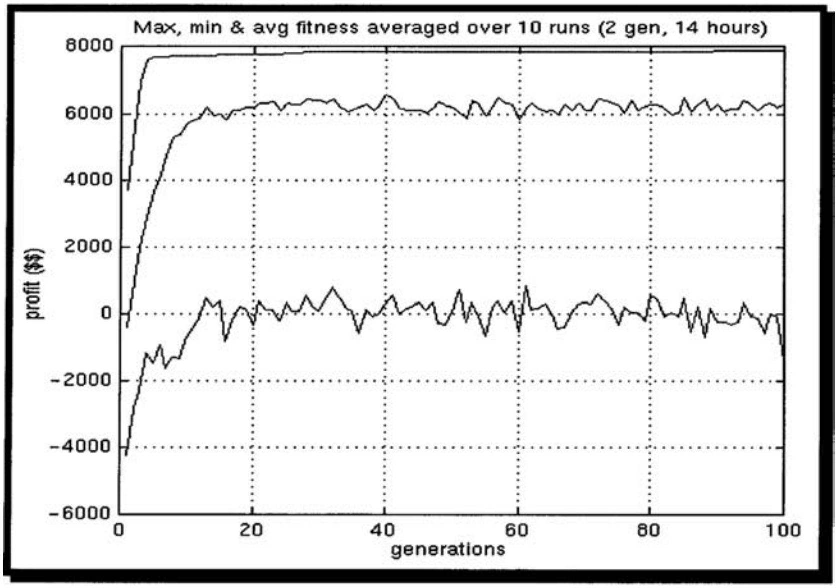


FIGURE 12.16 Max., min., and avg. fitness vs. GA generations for the 2-generator, 14-hour case.

Unit Commitment and Auctions

Regardless of the market framework, the solution method, and who is performing the UC, an auction can model and achieve the optimal solution. As mentioned previously in the section on EDC, auctions (which come in many forms, e.g., Dutch, English, sealed, double-sided, single-sided, etc.) are used to match buyers with sellers and to achieve a price that is considered fair. An auction can be used to find the optimal allocation, and the unit commitment algorithm essentially performs the same allocation that an auction would. Suppose an auctioneer was to call out a price, or a set of prices that is predicted for the schedule period. The auctioneer would then ask all generators how much power they would generate at that level. The generator must consider which units to switch on, and at what level to produce and sell. The reply amounts could be summed to determine the production level at that price. If all of the constraints, including demand, are met, then the most economical combination of units operating at the most economical settings has been found. If not, the auctioneer adjusts the price and asks for the amounts at the new price. This procedure is repeated until the constraints are satisfied. Prices may ascend as in the English auction, or they may descend as in the Dutch auction. See Fig. 12.7 for a graphical depiction of this process. For further discussion on this topic, the interested reader is referred to Sheblé (1999).

Summary of Economical Generation Operation

Since the introduction of electricity supply to the public in the late 1800s, people in many parts of the world have grown to expect an inexpensive reliable source of electricity. Providing that electric energy economically and efficiently requires the generation company to carefully control their generating units, and to consider many factors that may affect the performance, cost, and profitability of their operation. The unit commitment and economic dispatch algorithms play an important part in deciding how to operate the electric generating units around the world. The introduction of competition has changed many of the factors considered in solving these problems. Furthermore, advancements in solution techniques offer a continuum of candidate algorithms, each having its own advantages and disadvantages. Research continues to push these algorithms further. This chapter has provided the reader with an introduction to the problems of determining optimal unit commitment schedules and economic dispatches. It is by no means exhaustive, and the interested reader is strongly encouraged to see the references at the end of the chapter for more details.

References

Goldberg, D., *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Publishing Company, Inc., Reading, MA, 1989.

Kazarlis, S. A., Bakirtzis, A. G., and Petridis, V., A Genetic Algorithm Solution to the Unit Commitment Problem, *1995 IEEE/PES Winter Meeting,* 152-9 PWRS, New York, 1995.

Kondragunta, S., Genetic algorithm unit commitment program, M.S. Thesis, Iowa State University, Ames, IA, 1997.

Maifeld, T., and Sheblé, G., Genetic-Based unit commitment, *IEEE Trans. on Power Syst.,* 11, 1359, August 1996.

Richter, C., and Sheblé, G., A Profit-Based Unit Commitment GA for the Competitive Environment, accepted for *IEEE Trans. on Power Syst.,* publication forthcoming.

Sheblé, G., *Computational Auction Mechanisms for Restructured Power Industry Operation*. Kluwer Academic Publishers, Boston, MA, 1999.

Sheblé, G., Unit Commitment for Operations, Ph.D. Dissertation, Virginia Polytechnic Institute and State University, March, 1985.

Sheblé, G., and Fahd, G., Unit commitment literature synopsis, *IEEE Trans. on Power Syst.,* 9, 128–135, February 1994.

Takriti, S., Krasenbrink, B., and Wu, L. S.-Y., Incorporating Fuel Constraints and Electricity Spot Prices into the Stochastic Unit Commitment Problem, IBM Research Report: RC 21066, Mathematical Sciences Department, T.J. Watson Research Center, Yorktown Heights, New York, December 29, 1997.

Walters, D. C., and Sheblé, G. B., Genetic Algorithm Solution of Economic Dispatch with Valve Point Loading, *1992 IEEE/PES Summer Meeting,* 414-3, New York, 1992.

Wood, A., and Wollenberg, B., *Power Generation, Operation, and Control*. John Wiley & Sons, New York, NY, 1984.

12.3 State Estimation

Danny Julian

An online AC power flow is a valuable application when determining the critical elements affecting power system operation and control such as overloaded lines, credible contingencies, and unsatisfactory voltages. It is the basis for any real-time security assessment and enhancement applications.

AC power flow algorithms calculate real and reactive line flows based on a multitude of inputs with generator bus voltages, real power bus injections, and reactive power bus injections being a partial list. This implies that in order to calculate the line flows using a power flow algorithm, all of the input information (voltages, real power injections, reactive power injections, etc.) must be known *a priori* to the algorithm being executed.

An obvious way to implement an online AC power flow is to telemeter the required input information at every location in the power system. This would require not only a large number of **remote terminal units** (RTUs), but also an extensive communication infrastructure to telemeter the data to the **SCADA** system, both of which are costly. Although the generator bus voltages are usually readily available, the injection data is frequently what is lacking. This is because it is much easier and cheaper to monitor the net injection at a bus than to measure separate injections directly.

Also, this approach presents weaknesses for the online AC power flow that are due to meter accuracy and communication failure. An online power flow relying on a specific set of measurements could become unusable or give erroneous results if any of the predefined measurements became unavailable due to communication failure or due to misoperation of measurement devices. This is not a desirable outcome of an online application designed to alert system operators to unsecure conditions.

Given the above obstacles of utilizing an online AC power flow, work was conducted in the late 1960s and early 1970s (Schweppe and Wildes, Jan. 1970) into developing a process of performing an online power flow using not just the limited data needed for the classical AC power flow algorithm, but using all available measurements. This work led to the **state estimator**, which uses not only the aforementioned voltages but other telemetered measurements such as real and reactive line flows, circuit breaker statuses, and transformer tap settings.

State Estimation Problem

State estimators perform a statistical analysis using a set of m imperfect redundant data telemetered from the power system to determine the state of the system. The state of the system is a function of n **state variables**: bus voltages and relative phase angles, and tap changing transformer positions. Although the state estimation solution is not a "true" representation of the system, it is the "best" possible representation based on the telemetered measurements.

Also, it is necessary to have the number of measurements greater than the number of states ($m \geq n$) to yield a representation of the complete state of the system. This is known as the observability criterion. Typically, m is two to three times the value of n, allowing for a considerable amount of redundancy in the measurement set.

Underlying Assumptions

Telemetered measurements usually are corrupted since they are susceptible to noise. Even when great care is taken to ensure accuracy, unavoidable random noise enters into the measurement process, which distorts the telemetered values.

Fortunately, statistical properties associated with the measurements allow certain assumptions to be made to estimate the true measured value. First, it is assumed the measurement noise has an expected value, or average, of zero. This assumption implies the error in each measurement has equal probability of taking on a positive or negative value. It is also assumed that the expected value for the square of the measurement error is normal and has a standard deviation of σ, and the correlation between measurements is zero (i.e., independent).[1] A variable is said to be normal (or Gaussian) if its probability density function has the form

$$f(v) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{v^2}{2\sigma^2}}. \qquad (12.8)$$

This distribution is also known as the bell curve due to its symmetrical shape resembling a bell as can be seen in Fig. 12.17. The normal distribution is used for the modeling of measurement errors since it is the distribution that results when many factors contribute to the overall error.

Figure 12.17 also illustrates the effect of standard deviation on the normal density function. Standard deviation, σ, is a measure of the spread of the normal distribution about the mean (μ) and gives an indication of how many samples fall within a given interval around the mean. A large standard deviation implies there is a high probability the measurement noise will take on large values. Conversely, a small standard deviation implies there is a high probability the measurement noise will take on small values.

Measurement Representations

Since a measurement is not exact, it can be expressed with an error component of the form

$$z = z_T + v \qquad (12.9)$$

[1]In practice, measurements i and j are not necessarily independent since one measurement device may measure more than one value. Therefore, if the measurement device is bad, probably both measurements i and j are bad also.

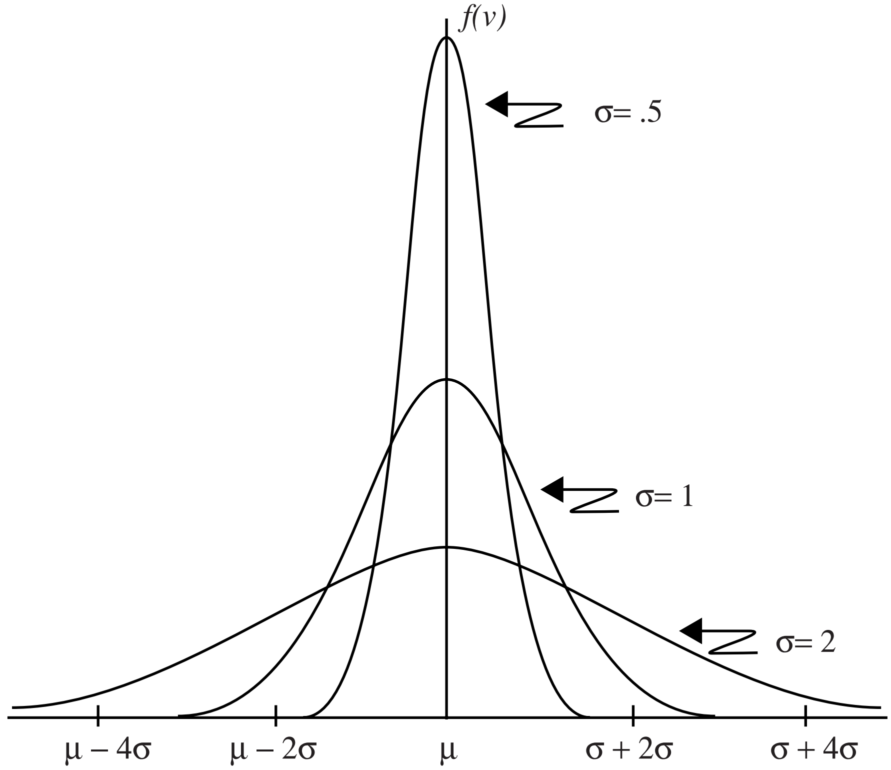


FIGURE 12.17 Normal probability distribution curve with a mean of μ.

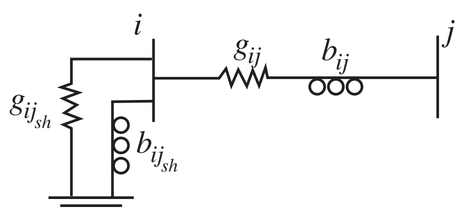


FIGURE 12.18 Transmission line representation.

where z is the measured value, z_T is the true value, and v is the measurement error that represents uncertainty in the measurement. In general, the measured value, as expressed in Eq. (12.9), can be related to the states, x, by

$$z = h(x) + v \tag{12.10}$$

where $h(x)$ is a vector of nonlinear functions relating the measurements to the state variables. An example of the $h(x)$ vector can be shown using the transmission line in Fig. 12.18.

Assuming real and reactive power measurements are being made at bus i in Fig. 12.18, the equations for line flow from bus i to j need to be determined as

$$P_{ij} = \left|\tilde{V}_i\right|^2 \left(g_{ij} + g_{i_{sh}}\right) - \left|\tilde{V}_i\right|\left|\tilde{V}_j\right| \left[g_{ij}\cos\left(\delta_{ij}\right) + b_{ij}\sin\left(\delta_{ij}\right)\right] \tag{12.11}$$

$$Q_{ij} = \left|\tilde{V}_i\right|^2 \left(b_{ij} + b_{i_{sh}}\right) - \left|\tilde{V}_i\right|\left|\tilde{V}_j\right| \left[g_{ij} \sin\left(\delta_{ij}\right) + b_{ij} \cos\left(\delta_{ij}\right)\right] \tag{12.12}$$

where $|\tilde{V}_i|$ is the magnitude of the voltage at bus i, $|\tilde{V}_j|$ is the magnitude of the voltage at bus j, δ_{ij} is the phase angle difference between bus i and bus j, g_{ij} and b_{ij} are the conductance and susceptance of line i-j, respectively, and $g_{i_{sh}}$ and $b_{i_{sh}}$ are the shunt conductance and susceptance at bus i, respectively.

Using Eqs. (12.11) and (12.12), Eq. (12.10) can now be rewritten as[2]

$$\begin{aligned} \bar{z} &= \bar{h}\left(x\right) + \bar{v} \\ &= \begin{bmatrix} \left|\tilde{V}_i\right|^2 \left(g_{ij} + g_{i_{sh}}\right) - \left|\tilde{V}_i\right|\left|\tilde{V}_j\right| \left[g_{ij} \cos\left(\delta_{ij}\right) + b_{ij} \sin\left(\delta_{ij}\right)\right] \\ \left|\tilde{V}_i\right|^2 \left(b_{ij} + b_{i_{sh}}\right) - \left|\tilde{V}_i\right|\left|\tilde{V}_i\right| \left[g_{ij} \sin\left(\delta_{ij}\right) + b_{ij} \cos\left(\delta_{ij}\right)\right] \end{bmatrix} + \begin{bmatrix} v_{P_{ij}} \\ v_{Q_{ij}} \end{bmatrix} \end{aligned} \tag{12.13}$$

which expresses the measurements entirely in terms of network parameters (which are assumed known) and system states (bus voltage and phase angle).

Solution Methods

The solution to the state estimation problem has been addressed by a broad class of techniques (Filho et al., Aug. 1990) and differs from power flow algorithms in two modes:

1. certain input data are either missing or inexact, and/or
2. the algorithm used for the calculation may entail approximations and approximate methods designed for high speed processing in the online environment.

In this section, two different solution methods to the state estimation problem will be introduced and described.

Weighted Least Squares

The most common approach to solving the state estimation problem is using the method of weighted least squares (WLS). This is accomplished by identifying the values of the state variables that minimize the performance index, J (the weighted sum of square errors):

$$J = \bar{e}^{\,T} R^{-1} \bar{e} \tag{12.14}$$

where the weighting factor, R, is the diagonal covariance matrix of the measurements and is defined as

$$E\left[\bar{v}\bar{v}^{\,T}\right] = R = \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \sigma_m^2 \end{bmatrix}. \tag{12.15}$$

By defining the error, e, in Eq. (12.14) as the difference between the true measured value, z, and the estimated measured value, $\hat{z}$,

$$\bar{e} = \bar{z} - \hat{\bar{z}} \tag{12.16}$$

[2]The superscript ‾ represents a vector.

a new form for the performance index can be written as

$$J = \left(\bar{z} - \bar{h}(x)\right)^T R^{-1}\left(\bar{z} - \bar{h}(x)\right) \tag{12.17}$$

As shown in Eqs. (12.15) and (12.17), the weights are defined by the inverse of the measurements variances. As a result, measurements of a higher quality have smaller variances that correspond to their weights having higher values, while measurements with poor quality have smaller weights due to the correspondingly higher variance values.

In order to minimize the performance index, J, a first-order necessary condition must hold, namely:

$$\left.\frac{\partial J}{\partial \bar{x}}\right|_{x^k} = 0 \tag{12.18}$$

Evaluating Eq. (12.17) at the necessary condition gives the following:

$$H\left(x^k\right)^T R^{-1}\left(\bar{z} - \bar{h}(x)\right) = 0 \tag{12.19}$$

where $H(x)$ represents the $m \times n$[3] measurement Jacobian matrix evaluated at iteration k:

$$H(x) = \left.\begin{bmatrix} \frac{\partial h_1}{\partial x_1} & \frac{\partial h_1}{\partial x_2} & \cdots & \frac{\partial h_1}{\partial x_n} \\ \frac{\partial h_2}{\partial x_1} & \frac{\partial h_2}{\partial x_2} & \cdots & \frac{\partial h_2}{\partial x_n} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial h_m}{\partial x_1} & \frac{\partial h_m}{\partial x_2} & \cdots & \frac{\partial h_m}{\partial x_n} \end{bmatrix}\right|_{x^k} \tag{12.20}$$

A linearized relationship between the measurements and the state variables is then found by expanding the Taylor series expansion of the function $\bar{h}(x)$ around a point x^k:

$$\bar{h}\left(x^k\right) = \bar{h}\left(x^k\right) + \Delta\bar{x}^k \frac{\partial \bar{h}\left(x^k\right)}{\partial \bar{x}} + \text{higher order terms.} \tag{12.21}$$

This set of equations can be solved using an iterative approach such as Newton Raphson's method. At the (k+1)th iteration, the refreshed values of the state variables can be obtained from their values in the previous iteration by:

$$\bar{x}^{k+1} = \bar{x}^k + \left(H\left(x^k\right)^T R^{-1} H\left(x^k\right)\right)^{-1} H\left(x^k\right)^T R^{-1}\left(\bar{z} - \bar{h}\left(x^k\right)\right). \tag{12.22}$$

At convergence, the solution $\bar{x}^{k+1}$ corresponds to the weighted least squares estimates of the state variables. Convergence can be determined either by satisfying

[3] m represents the number of measurements; n represents the number of states.

$$\max\left(\bar{x}^{k+1} - \bar{x}^{k}\right) \le \varepsilon \tag{12.23}$$

or

$$J^{k+1} - J^{k} \le \varepsilon \tag{12.24}$$

where ε is some predetermined convergence factor.

Linear Programming

Another solution method that addresses the state estimation problem is linear programming. Linear programming is an optimization technique that serves to minimize a linear objective function subject to a set of constraints:

$$\begin{aligned} &\min\left\{\bar{c}^{T}\bar{x}\right\} \\ &s.t. \quad A\bar{\mathrm{x}} = \bar{\mathrm{b}} \\ &\qquad \bar{x} \ge 0 \end{aligned} \tag{12.25}$$

There are many different techniques associated with solving linear programming problems including the simplex and interior point methods.

Since the objective function, as expressed in Eq. (12.17), is quadratic in terms of the unknowns (states), it must be rewritten in a linear form. This is accomplished by first rewriting the measurement error, as expressed in Eq. (12.10), in terms of a positive measurement error, v_p, and a negative measurement error, v_n:

$$\begin{aligned} \bar{z} &= \bar{h}(x) + \bar{v} \\ &= \bar{h}(x) + \bar{v}_p - \bar{v}_n \end{aligned} \tag{12.26}$$

Restricting the positive and negative measurement errors to only nonnegative values insures the problem is bounded. This was not a concern in the weighted least squares approach since a quadratic function is convex and is guaranteed to contain a global minimum.

Using the new definition of a measurement described in Eq. (12.26) and the inverse of the diagonal covariance matrix of the measurements for weights as described in the weighted least squares approach, the objective function can now be written as:

$$J = R^{-1}\left(\bar{v}_p + \bar{v}_n\right) \tag{12.27}$$

The constraints are the equations relating the state vector to the measurements as shown in Eq. (12.26). Once again, since $h(x)$ is nonlinear, it must be linearized around a point x^k by expanding the Taylor series, as was performed previously in the weighted least squares approach. The solution to the state estimation problem can then be determined by solving the following linear program:

$$\begin{aligned} &\min\left\{R^{-1}\left(\bar{v}_p + \bar{v}_n\right)\right\} \\ &s.t. \quad \Delta\bar{\mathrm{z}}^{\mathrm{k}} - \mathrm{H}\left(\mathrm{x}^{\mathrm{k}}\right)\Delta\bar{x}^{k} + \bar{v}_p - \bar{v}_n = 0 \\ &\qquad \bar{v}_p \ge 0 \\ &\qquad \bar{v}_n \ge 0 \end{aligned} \tag{12.28}$$

where $H(x^k)$ represents the $m \times n$ measurement Jacobian matrix evaluated at iteration k as defined in Eq. (12.20).

State Estimation Operation

State estimators are typically executed either periodically (i.e., every 5 min), on demand, or due to a status change such as a breaker operation isolating a line section. To illustrate the relationship of the state estimator with respect to other **EMS** applications, a simple depiction of an EMS is shown below in Fig. 12.19:

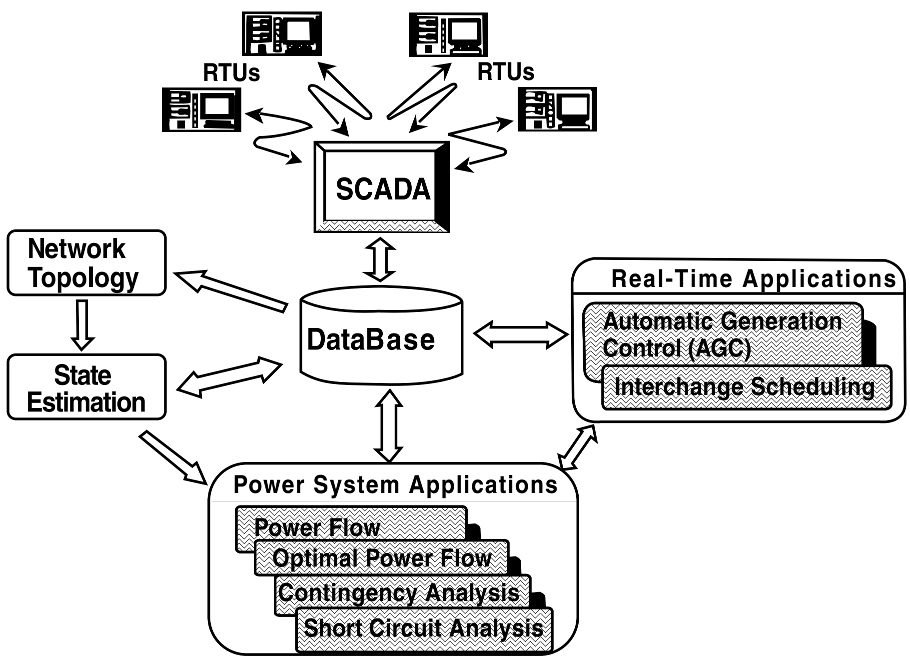


FIGURE 12.19 Simple depiction of an EMS.

As shown, the state estimator receives inputs from the supervisory control and data acquisition (SCADA) system and the network topology assessment applications and stores the state of the system in a central location (i.e., database). Power system applications, such as contingency analysis and optimal power flow, can then be executed based on the state of the system as computed by the state estimator.

Network Topology Assessment

Before the state estimator is executed in realtime, the topology of the network is determined. This is accomplished by a system or **network configurator** that establishes the configuration of the power system network based on telemetered breaker and switch statuses. The network configurator normally addresses questions like:

- Have breaker operations caused individual buses to either be split into two or more isolated buses, or combined into a single bus?
- Have lines been opened or restored to service?

The state estimator then uses the network determined by the network configurator, which consists only of energized (online) lines and devices, as a basis for the calculations to determine the state of the system.

Error Identification

Since state estimators utilize telemetered measurements and network parameters as a foundation for their calculations, the performance of the state estimator depends on the accuracy of the measured data as well as the parameters of the network model. Fortunately, the use of all available measurements introduces a favorable secondary effect caused by the redundancy of information. This redundancy provides the state estimator with more capabilities than just an online AC power flow; it introduces the ability to detect "bad" data. Bad data can come from many sources, such as:

- approximations,
- simplified model assumptions,
- human data handling errors, or
- measurement errors due to faulty devices (e.g., transducers, current transformers).

Telemetered Data

The ability to detect and identify bad measurements is an extremely useful feature of the state estimator. Without the state estimator, obviously wrong telemetered measurements would have little chance of being identified. With the state estimator, operation personnel can have a greater confidence that telemetered data is not grossly in error.

Data is tagged as "bad" when the estimated value is unreasonably different from the measured/telemetered value obtained from the RTU. As a simple example, suppose a bus voltage is measured to be 1.85 pu and is estimated to be 0.95 pu. In this case, the bus voltage measurement could be tagged as bad. Once data is tagged as bad, it should be removed from the measurement set before being utilized by the state estimator.

Most state estimators rely on a combination of preestimation and postestimation schemes for detection and elimination of bad data. Preestimation involves gross bad data detection and consistency tests. Data is identified as bad in preestimation by the detection of gross measurement errors such as zero voltages or line flows that are outside reasonable limits using network topology assessment. Consistency tests classify data either as valid, suspect, or raw for use in postestimation analysis by using statistical properties of related measurements. Measurements are classified as valid if they pass a consistency test that separates measurements into subsets based on a consistency threshold. If the measurement fails the consistency test, it is classified as suspect. Measurements are classified as raw if a consistency test cannot be made and they cannot be grouped into any subset. Raw measurements typically belong to nonredundant portions of the complete measurement set.

Postestimation involves performing a statistical analysis (e.g., hypothesis testing using chi-square tests) on the normalized measurement residuals. A normalized residual is defined as

$$r_i = \frac{z_i - h_i(x)}{\sigma_i} \tag{12.29}$$

where σ_i is the i-th diagonal term of the covariance matrix, R, as defined in Eq. (12.15). Data is identified as bad in postestimation typically when the normalized residuals of measurements classified as suspect lie outside a predefined confidence interval (i.e., fail the chi-square test).

Parameter Data

In parameter error identification, network parameters (i.e., admittances) that are suspicious are identified and need to be estimated. The use of faulty network parameters can severely impact the quality of state estimation solutions and cause considerable error. A requirement for parameter estimation is that all parameters be identifiable by measurements. This requirement implies the lines under consideration have associated measurements, thereby increasing the size of the measurement set by l, where l is the number of parameters to be estimated. Therefore, if parameter estimation is to be performed, the observability criterion must be augmented to become $m \geq n + l$.

Unobservability

By definition, a state variable is unobservable if it cannot be estimated. Unobservability occurs when the observability criterion is violated ($m < n$) and there are insufficient redundant measurements to determine the state of the system. Mathematically, the matrix $H(x^k)^T R^{-1} H(x^k)$ of Eq. (12.22) becomes singular and cannot be inverted.

The obvious solution to the unobservability problem is to increase the number of measurements. The problem then becomes where and how many measurements need to be added to the measurement set. Adding additional measurements is costly since there are many supplementary factors that must be addressed in addition to the cost of the measuring device such as RTUs, communication infrastructure, and software data processing at the EMS. A number of approaches have been suggested that try to minimize the cost while satisfying the observability criterion (Baran et al., Aug. 1995; Park et al., Aug. 1998).

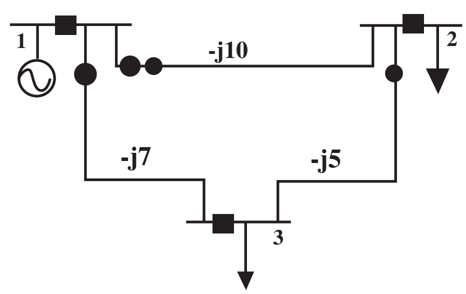
Another solution to address the problem of unobservability is to augment the measurement set with pseudomeasurements to reach an observability condition for the network. When adding pseudomeasurements to a network, the equation of the pseudomeasured quantity is substituted for actual measurements. In this case, the measurement covariance values in Eq. (12.15) associated with these measurements should have large values that allow the state estimator to treat the pseudomeasurements as if they were measured from a very poor metering device.

Example State Estimation Problem

This section provides a simple example to illustrate how the state estimation process is performed. The WLS method, as previously described, will be applied to a sample system.

System Description

A sample three-bus system is shown below in Fig. 12.20:



where:

■ ⇒ Voltage Measurement (V)
● ⇒ Real Power Measurement (MW)
• ⇒ Reactive Power Measurement (MVAr)

FIGURE 12.20 Sample three-bus power flow system.

Bus 1 is assumed to be the reference bus with a corresponding angle of zero. All other relevant system data is given in Table 12.7.

WLS State Estimation Process

First, the states (x) are defined as the angles at bus 2 and bus 3 and the voltage magnitudes at all buses[4]:

[4] The angle at bus one is not chosen as a state since it is designated as the reference bus.

TABLE 12.7 Sample System Data

Measurement Type	Measurement Location	Measurement Value (pu)	Measurement Covariance (σ)
$\lvert\tilde{V}\rvert$	Bus 1	1.02	0.05
$\lvert\tilde{V}\rvert$	Bus 2	1.0	0.05
$\lvert\tilde{V}\rvert$	Bus 3	0.99	0.05
P	Bus 1 – Bus 2	1.5	0.1
Q	Bus 1 – Bus 2	0.2	0.1
P	Bus 1 – Bus 3	1.0	0.1
Q	Bus 2 – Bus 3	0.1	0.1

$$\bar{x} = \begin{bmatrix} \delta_2 \\ \delta_3 \\ \lvert\tilde{V}_1\rvert \\ \lvert\tilde{V}_2\rvert \\ \lvert\tilde{V}_3\rvert \end{bmatrix}$$

This gives a total of seven measurements and five states that satisfy the observability criterion requiring more measurements than states.

Using the previously defined equations for the WLS state estimation procedure, the following can be determined:

$$R = \left[\sigma_i^2\right]$$

$$= \begin{bmatrix} (.05)^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & (.05)^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & (.05)^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & (.1)^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & (.1)^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & (.1)^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & (.1)^2 \end{bmatrix}$$

$$\hat{\bar{z}} = h(\bar{x})$$

$$= \begin{bmatrix} x_3 \\ x_4 \\ x_5 \\ 10x_3x_4 \sin x_1 \\ 10x_3^2 - 10x_3x_4 \cos x_1 \\ 7x_3x_5 \sin x_2 \\ 5x_4^2 - 5x_4x_5 \cos(x_1 - x_2) \end{bmatrix}$$

$$H(x) = \left[\frac{\partial \bar{h}}{\partial \bar{x}}\right]$$

$$= \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 10x_3x_4 \cos x_1 & 0 & 10x_4 \sin x_1 & 10x_3 \sin x_1 & 0 \\ 10x_3x_4 \sin x_1 & 0 & 20x_3 - 10x_4 \cos x_1 & -10x_3 \cos x_1 & 0 \\ 0 & 7x_3x_5 \cos x_2 & 7x_5 \sin x_2 & 0 & 7x_3 \sin x_2 \\ 5x_4x_5 \sin(x_1 - x_2) & -5x_4x_5 \sin(x_1 - x_2) & 0 & 10x_4 - 5x_5 \cos(x_1 - x_2) & -5x_4 \cos(x_1 - x_2) \end{bmatrix}$$

Using zero as an initial guess for the states representing voltage angles (x_1 and x_2) and the measured voltages as given in Table 12.7 for the states representing voltage magnitudes (x_3, x_4, and x_5):

$$\begin{bmatrix} x_1^0 \\ x_2^0 \\ x_3^0 \\ x_4^0 \\ x_5^0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1.02 \\ 1.00 \\ 0.99 \end{bmatrix},$$

the state values at the first iteration are determined by Eq. (12.22) to be

$$\begin{bmatrix} x_1^1 \\ x_2^1 \\ x_3^1 \\ x_4^1 \\ x_5^1 \end{bmatrix} = \begin{bmatrix} 0.147 \\ 0.142 \\ 1.022 \\ 1.003 \\ 0.984 \end{bmatrix}.$$

After four iterations, the state estimation process converges to the final states:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 0.147 \\ 0.143 \\ 1.016 \\ 1.007 \\ 0.987 \end{bmatrix}.$$

Using the solved voltages and angles from the state estimation process, the line flows and bus injections can now be calculated. With the state of the system now known, other applications such as contingency analysis and optimal power flow may be performed. Notice, the state estimation process results in the state of the system, just as when performing a power flow but without *a priori* knowledge of bus injections.

Defining Terms

Remote Terminal Unit (RTU): Hardware that telemeters systemwide data from various field locations (i.e., substations, generating plants) to a central location.

State estimator: An application that uses a statistical process in order to estimate the state of the system.

State variable: The quantity to be estimated by the state estimator, typically bus voltage and angle.

Network configurator: An application that determines the configuration of the power system based on telemetered breaker and switch statuses.

Supervisory Control and Data Acquisition (SCADA): A computer system that performs data acquisition and remote control of a power system.

Energy Management System (EMS): A computer system that monitors, controls, and optimizes the transmission and generation facilities with advanced applications. A SCADA system is a subset of an EMS.

References

Schweppe, F.C., Wildes, J., Power System Static-State Estimation I,II,III, *IEEE Trans. on Power Appar. Syst.,* 89, 120–135, January 1970.

Filho, M.B.D.C. et al., Bibliography on power system state estimation (1968-1989), *IEEE Trans. on Power Syst.,* 5, 3, 950–961, August 1990.

Baran, M.E. et al., A meter placement method for state estimation, *IEEE Trans. on Power Syst.,* 10, 3, 1704–1710, August 1995.

Park, Y.M et al., Design of reliable measurement system for state estimation, *IEEE Trans. on Power Syst.,* 3, 3, 830–836, August 1998.

12.4 Optimal Power Flow

M. E. El-Hawary

An Optimal Power Flow (OPF) function schedules the power system controls to optimize an objective function while satisfying a set of nonlinear equality and inequality constraints. The equality constraints are the conventional power flow equations; the inequality constraints are the limits on the control and operating variables of the system. Mathematically, the OPF can be formulated as a constrained nonlinear optimization problem. This section reviews features of the problem and some of its variants as well as requirements for online implementation.

Optimal scheduling of the operations of electric power systems is a major activity, which turns out to be a large-scale problem when the constraints of the electric network are taken into account. This document deals with recent developments in the area emphasizing optimal power flow formulation and deals with conventional optimal power flow (OPF), accounting for the dependence of the power demand on voltages in the system, and requirements for online implementation.

The OPF problem was defined in the early 1960s (Burchett et al., Feb. 1982) as an extension of conventional economic dispatch to determine the optimal settings for control variables in a power network respecting various constraints. OPF is a static constrained nonlinear optimization problem, whose development has closely followed advances in numerical optimization techniques and computer technology. It has since been generalized to include many other problems. Optimization of the electric system with losses represented by the power flow equations was introduced in the 1960s (Carpentier, 1962; Dommel and Tinney, Oct. 1968). Since then, significant effort has been spent on achieving faster and robust solution methods that are suited for online implementation, operating practice, and security requirements.

OPF seeks to optimize a certain objective, subject to the network power flow constraints and system and equipment operating limits. Today, any problem that involves the determination of the instantaneous "optimal" steady state of an electric power system is referred to as an Optimal Power Flow problem. The optimal steady state is attained by adjusting the available controls to minimize an objective function subject to specified operating and security requirements. Different classes of OPF problems, designed for special-purpose applications, are created by selecting different functions to be minimized, different sets of controls, and different sets of constraints. All these classes of the OPF problem are subsets of the general problem. Historically, different solution approaches have been developed to solve these different classes of OPF. Commercially available OPF software can solve very large and complex formulations in a relatively short time, but may still be incapable of dealing with online implementation requirements.

There are many possible objectives for an OPF. Some commonly implemented objectives are:

- fuel or active power cost optimization,
- active power loss minimization,
- minimum control-shift,
- minimum voltage deviations from unity, and
- minimum number of controls rescheduled.

In fuel cost minimization, the outputs of all generators, their voltages, LTC transformer taps and LTC phase shifter angles, and switched capacitors and reactors are control variables. The active power losses can be minimized in at least two ways (Happ and Vierath, July, 1986). In both methods, all the above variables are adjusted except for the active power generation. In one method, the active power generation at the swing bus is minimized while keeping all other generation constant at prespecified values. This effectively minimizes the total active power losses. In another method, an actual expression for the losses is minimized, thus allowing the exclusion of lines in areas not optimized.

The behavior of the OPF solutions during contingencies was a major concern, and as a result, security constrained optimal power flow was introduced in the early 1970s. Subsequently, online implementations became a new thrust in order to meet the challenges of new deregulated operating environments.

Conventional Optimal Economic Scheduling

Conventional optimal economic scheduling minimizes the total fuel cost of thermal generation, which may be approximated by a variety of expressions such as linear or quadratic functions of the active power generation of the unit. The total active power generation in the system must equal the load plus the active transmission losses, which can be expressed by the celebrated Kron's loss formula. Reserve constraints may be modeled depending on system requirements. Area and system spinning, supplemental, emergency, or other types of reserve requirements involve functional inequality constraints. The forms of the functions used depend on the type of reserve modeled. A linear form is evidently most attractive from a solution method point of view. However, for thermal units, the spinning reserve model is nonlinear due to the limit on a unit's maximum reserve contribution. Additional constraints may be modeled, such as area interchange constraints used to model network transmission capacity limitations. This is usually represented as a constraint on the net interchange of each area with the rest of the system (i.e., in terms of limits on the difference between area total generation and load).

The objective function is augmented by the constraints using a Lagrange-type multiplier lambda, λ. The optimality conditions are made up of two sets. The first is the problem constraints. The second set is based on variational arguments giving for each thermal unit:

$$\frac{\partial F_i}{\partial P_i} = \lambda \left[1 - \frac{\partial P_L}{\partial P_i} \right] \quad \mathrm{i} = 1, \ldots, \mathrm{N} \tag{12.30}$$

The optimality conditions along with the physical constraints are a set of nonlinear equations that requires iterative methods to solve. Newton's method has been widely accepted in the power industry as a powerful tool to solve problems such as the load flow and optimal load flow. This is due to its reliable and fast convergence, known to be quadratic.

A solution can usually be obtained within a few iterations, provided that a reasonably good initial estimate of the solution is available. It is therefore appropriate to employ this method to solve the present problem.

Conventional OPF Formulation

The optimal power flow is a constrained optimization problem requiring the minimization of:

$$f = (\boldsymbol{x}, \boldsymbol{u}) \tag{12.31}$$

subject to

$$\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{u}) = 0 \tag{12.32}$$

$$\boldsymbol{h}(\boldsymbol{x}, \boldsymbol{u}) \le 0 \tag{12.33}$$

$$\boldsymbol{u}^{\min} \le \boldsymbol{u} \le \boldsymbol{u}^{\max} \tag{12.34}$$

$$\boldsymbol{x}^{\min} \le \boldsymbol{x} \le \boldsymbol{x}^{\max} \tag{12.35}$$

Here **f**(**x**,**u**) is the scalar objective function, **g**(**x**,**u**) represents nonlinear equality constraints (power flow equations), and **h**(**x**,**u**) is the nonlinear inequality constraint of vector arguments **x** and **u**. The vector **x** contains dependent variables consisting of bus voltage magnitudes and phase angles, as well as the MVAr output of generators designated for bus voltage control and fixed parameters such as the reference bus angle, noncontrolled generator MW and MVAr outputs, noncontrolled MW and MVAr loads, fixed bus voltages, line parameters, etc. The vector **u** consists of control variables including:

- real and reactive power generation
- phase-shifter angles
- net interchange
- load MW and MVAr (load shedding)
- DC transmission line flows
- control voltage settings
- LTC transformer tap settings

Examples of equality and inequality constraints are:

- limits on all control variables
- power flow equations
- generation/load balance
- branch flow limits (MW, MVAr, MVA)
- bus voltage limits
- active/reactive reserve limits
- generator MVAr limits
- corridor (transmission interface) limits

The power system consists of a total of N buses, N_G of which are generator buses. M buses are voltage controlled, including both generator buses and buses at which the voltages are to be held constant. The voltages at the remaining (N – M) buses (load buses), must be found.

The network equality constraints are represented by the load flow equations:

$$P_i(V, \delta) - P_{gi} + P_{di} = 0 \tag{12.36}$$

$$Q_i(V, \delta) - Q_{gi} + Q_{di} = 0 \tag{12.37}$$

Two different formulation versions can be considered.

(a) *Polar Form:*

$$P_i(V,\delta) = |V_i| \sum_{1}^{N} |V_i||Y_{ij}| \cos(\delta_i - \delta_j - \phi_{ij}) \tag{12.38}$$

$$Q_i(V,\delta) = |V_i| \sum_{1}^{N} |V_j||Y_{ij}| \sin(\delta_i - \delta_j - \phi_{ij}) \tag{12.39}$$

$$Y_{ij} = |Y_{ij}| \angle \varphi_{ij} \tag{12.40}$$

where

P_i = Active power injection into bus i.
Q_i = Reactive power injection into bus i.
$|\tilde{V}_i|$ = Voltage magnitude of bus i.
δ_i = Angle at bus i.
$|\tilde{\mathbf{Y}}_{ij}|$, φ_{ij} = Magnitude and angle of the admittance matrix.
P_{di}, Q_{di} = Active and reactive load on bus i.

(b) *Rectangular Form:*

$$P_i(e,f) = e_i\left[\sum_{1}^{N}(G_{ij}e_j - B_{ij}f_j)\right] + f_i\left[\sum_{1}^{N}(G_{ij}f_j + B_{ij}e_j)\right] \tag{12.41}$$

$$Q_i(e,f) = f_i\left[\sum_{1}^{N}(G_{ij}e_j - B_{ij}f_j)\right] - e_i\left[\sum_{1}^{N}(G_{ij}f_j + B_{ij}e_j)\right] \tag{12.42}$$

e_i = Real part of complex voltage at bus i.
f_i = Imaginary part of the complex voltage at bus i.
G_{ij} = Real part of the complex admittance matrix.
B_{ij} = Imaginary part of the complex admittance matrix.

The control variables vary according to the objective being minimized. For fuel cost minimization, they are usually the generator voltage magnitudes, generator active powers, and transformer tap ratios. The dependent variables are the voltage magnitudes at load buses, phase angles, and reactive generations.

Application of Optimization Methods to OPF

Various optimization methods have been proposed to solve the optimal power flow problem, some of which are refinements on earlier methods. These include:

1. Generalized Reduced Gradient (GRG) method.
2. Reduced gradient method.
3. Conjugate gradient methods.
4. Hessian-based method.
5. Newton's method.
6. Linear programming methods.
7. Quadratic programming methods.
8. Interior point methods.

Some of these techniques have spawned production OPF programs that have achieved a fair level of maturity and have overcome some of the earlier limitations in terms of flexibility, reliability, and performance requirements.

Generalized Reduced Gradient Method

The Generalized Reduced Gradient method (GRG), due to Abadie and Carpentier (1969), is an extension of the Wolfe's reduced gradient method (Wolfe, 1967) to the case of nonlinear constraints. Peschon in 1971 and Carpentier in 1973 used this method for OPF. Others have used this method to solve the optimal power flow problem since then (Lindqvist et al., 1984; Yu et al., 1986).

Reduced Gradient Method

A reduced gradient method was used by Dommel and Tinney (1968). An augmented Lagrangian function is formed. The negative of the gradient $\partial L/\partial u$ is the direction of steepest descent. The method of reduced gradient moves along this direction from one feasible point to another with a lower value of f, until the solution does not improve any further. At this point an optimum is found, if the Kuhn-Tucker conditions (1951) are satisfied. Dommel and Tinney used Newton's method to solve the power flow equations.

Conjugate Gradient Method

In 1982, Burchett et al. used a conjugate gradient method, which is an improvement on the reduced gradient method. Instead of using the negative gradient ∇f as the direction of steepest descent, the descent directions at adjacent points are linearly combined in a recursive manner.

$$\Gamma_k = -\nabla f + \beta_k \Gamma_{k-1} \qquad \beta_0 = 0 \tag{12.43}$$

Here, r_k is the descent direction at iteration "k."

Two popular methods for defining the scalar value β_k are the Fletcher-Reeves method (Carpentier, June 1973) and the Polak-Ribiere method (1969).

Hessian-Based Methods

Sasson (Oct. 1969) discusses methods (Fiacco and McCormick, 1964; Lootsma, 1967; Zangwill, 1967) that transform the constrained optimization problem into a sequence of unconstrained problems. He uses a transformation introduced by Powell and Fletcher (1963). Here, the Hessian matrix is not evaluated directly. Instead, it is built indirectly starting initially with the identity matrix so that at the optimum point it becomes the Hessian itself.

Due to drawbacks of the Fletcher-Powell method, Sasson et al. (1973) developed a Hessian load flow with an extension to OPF. Here, the Hessian is evaluated and solved unlike in the previous method. The objective function is transformed as before to an unconstrained objective. An unconstrained objective is formed. All equality constraints and only the violating inequality constraints are included. The sparse nature of the Hessian is used to reduce storage and computation time.

Newton OPF

Newton OPF has been formulated by Sun et al. (1984), and later by Maria et al. (Aug. 1987). An augmented Lagrangian is first formed. The set of first derivatives of the augmented objective with respect to the control variables gives a set of nonlinear equations as in the Dommel and Tinney method. Unlike in the Dommel and Tinney method where only a part of these are solved by the N-R method, here, all equations are solved simultaneously by the N-R method.

The method itself is quite straightforward. It is the method of identifying binding inequality constraints that challenged most researchers. Sun et al. use a multiply enforced, zig-zagging guarded technique for some of the inequalities, together with penalty factors for some others. Maria et al. used an LP-based technique to identify the binding inequality set. Another approach is to use purely penalty factors. Once the binding inequality set is known, the N-R method converges in a very few iterations.

Linear Programming-Based Methods

LP methods use a linear or piecewise-linear cost function. The dual simplex method is used in some applications (Bentall, 1968; Shen and Laughton, Nov. 1970; Stott and Hobson, Sept./Oct. 1978; Wells, 1968). The network power flow constraints are linearized by neglecting the losses and the reactive powers, to obtain the DC load flow equations. Merlin (1972) uses a successive linearization technique and repeated application of the dual simplex method.

Due to linearization, these methods have a very high speed of solution, and high reliability in the sense that an optimal solution can be obtained for most situations. However, one drawback is the inaccuracies of the linearized problem. Another drawback for loss minimization is that the loss linearization is not accurate.

Quadratic Programming Methods

In these methods, instead of solving the original problem, a sequence of quadratic problems that converge to the optimal solution of the original problem are solved. Burchett et al. use a sparse implementation of this method. The original problem is redefined as simply, to minimize,

$$f(x) \tag{12.44}$$

subject to:

$$g(x) = 0 \tag{12.45}$$

The problem is to minimize

$$g^T p + \frac{1}{2} p^T Hp \tag{12.46}$$

subject to:

$$Jp = 0 \tag{12.47}$$

where

$$p = x - x_k \tag{12.48}$$

Here, g is the gradient vector of the original objective function with respect to the set of variables "x." "J" is the Jacobian matrix that contains the derivatives of the original equality constraints with respect to the variables, and "H" is the Hessian containing the second derivatives of the objective function and a linear combination of the constraints with respect to the variables. x_k is the current point of linearization. The method is capable of handling problems with infeasible starting points and can also handle ill-conditioning due to poor R/X ratios. This method was later extended by El-Kady et al. (May 1986) in a study for the Ontario Hydro System for online voltage/var control. A nonsparse implementation of the problem was made by Glavitsch (Dec. 1983) and Contaxis (May, 1986).

Interior Point Methods

The projective scaling algorithm for linear programming proposed by N. Karmarkar is characterized by significant speed advantages for large problems reported to be as much as 50:1 when compared to the simplex method (Karmarkar, 1984). This method has a polynomial bound on worst-case running time that is better than the ellipsoid algorithms. Karmarkar's algorithm is significantly different from Dantzig's simplex method. Karmarkar's interior point rarely visits too many extreme points before an optimal point is found. The IP method stays in the interior of the polytope and tries to position a current solution as the "center of the universe" in finding a better direction for the next move. By properly choosing the step lengths, an optimal solution is achieved after a number of iterations. Although this IP approach requires more computational time in finding a moving direction than the traditional simplex method, better moving direction is achieved resulting in less iterations. Therefore, the IP approach has become a major rival of the simplex method and has attracted attention in the optimization community. Several variants of interior points have been proposed and successfully applied to optimal power flow (Momoh, 1992; Vargas et al., 1993; Yan and Quintana, 1999).

OPF Incorporating Load Models

Load Modeling

The area of power systems load modeling has been well explored in the last two decades of the twentieth century. Most of the work done in this area has dealt with issues in stability of the power system. Load modeling for use in power flow studies has been treated in a few cases (Concordia and Ihara, 1982; IEEE Committee Report, 1973; IEEE Working Group Report, 1996; Iliceto et al., 1972; Vaahedi et al., 1987). In stability studies, frequency and time are variables of interest, unlike in power flow and some OPF studies. Hence, load models for use in stability studies should account for any load variations with frequency and time as well. These types of load models are normally referred to as dynamic load models. In power flow, OPF studies neglecting contingencies, and security-constrained OPF studies using preventive control, time, and frequency, are not considered as variables. Hence, load models for this type of study need not account for time and frequency. These load models are static load models.

In security-constrained OPF studies using corrective control, the time allowed for certain control actions is included in the formulation. However, this time merely establishes the maximum allowable correction, and any dynamic behavior of loads will usually end before any control actions even begin to function. Hence, static load models can be used even in this type of formulation.

Static Load Models

Several forms of static load models have been proposed in the literature, from which the exponential and quadratic models are most commonly used. The exponential form is expressed as:

$$P_m = a_p V^{b_b} \tag{12.49}$$

$$Q_m = a_q V^{bq} \tag{12.50}$$

The values of the coefficients a_p and a_q can be taken as the specified active and reactive powers at that bus, provided the specified power demand values are known to occur at a voltage of 1.0 per unit, measured at the network side of the distribution transformer. A typical measured value of the demand and the network side voltage is sufficient to determine approximately the values of the coefficients, provided the exponents are known. The range of values reported for the exponents vary in the literature, but typical values are 1.5 and 2.0 for b_p and b_q, respectively.

Conventional OPF Studies Including Load Models

Incorporation of load models in OPF studies has been considered in a couple of cases (El-Din et al., 1989; Vaahedi and El-Din, May 1989) for the Ontario Hydro energy management system. In both cases, loss minimization was considered to be the objective. It is concluded by Vaahedi and El-Din (1989) that the modeling of ULTC operation and load characteristics is important in OPF calculations.

The effects of load modeling in OPF studies have been considered for the case where the generator bus voltages are held at prespecified values (Dias and El-Hawary, 1989). Since the swing bus voltage is held fixed at all times (and also the generator bus voltages in the absence of reactive power limit violations), the average system voltage is maintained in most cases. Thus, an increase in fuel cost due to load modeling was noticed for many systems that had a few (or zero) reactive limit violations, and a decrease for those with a noticeable number of reactive limit violations. Holding the generator bus voltages at specified values restricts the available degrees of freedom for OPF and makes the solution less optimal.

Incorporation of load models in OPF studies minimizing fuel cost (with all voltages free to vary within bounds) can give significantly different results when compared with standard OPF results. The reason for this is that the fuel cost can now be reduced by lowering the voltage at the modeled buses along with all other voltages wherever possible. The reduction of the voltages at the modeled buses lowers the power demand of the modeled loads and will thus give the lower fuel cost. When a large number of loads are modeled, the total fuel cost may be lower than the standard OPF. However, a lowering of the fuel cost via a lowering of the power demand may not be desirable under normal circumstances, as this will automatically decrease the total revenue of the operation. This can also give rise to a lower net revenue if the decrease in the total revenue is greater than the decrease in the fuel cost. This is even more undesirable. What is needed is an OPF solution that does not decrease the total power demand in order to achieve a minimum fuel cost. The standard OPF solution satisfies this criterion. However, given a fair number of loads that are fed by fixed tap transformers, the standard OPF solution can be significantly different from the practically observed version of this solution.

Before attempting to find an OPF solution incorporating load models that satisfies the required criterion, we deal with the reason for the problem. In a standard OPF formulation, the total revenue is constant and independent of the solution. Hence, we can define net revenue R_N, which is linearly related to the total fuel cost F_C by the formula:

$$R_N = -F_C + constant \tag{12.51}$$

The constant term is the total revenue dependent on the total power demand and the unit price of electricity charged to the customers. From this relationship we see that a solution with minimum fuel cost will automatically give maximum net revenue. Now, when load models are incorporated at some buses, the total power demand is not a constant, and hence the total revenue will also not be constant. As a result,

$$R_N = -F_C + R_T \tag{12.52}$$

where "R_T" is the total demand revenue and is no longer a constant.

If instead of minimizing the fuel cost, we now maximize the net revenue, we will definitely avoid the difficulties encountered earlier. This is equivalent to minimizing the difference between the fuel cost and the total revenue. Hence we see that, in the standard OPF, the required maximum net revenue is implied, and the equivalent minimum fuel cost is the only function that enters the computations.

Security Constrained OPF Including Load Models

A conventional OPF result can have optimal but insecure states during certain contingencies. This can be avoided by using a security constrained OPF. Unlike in the former, for a security constrained OPF, we can incorporate load models in a variety of ways. For example, we can consider the loads as independent of voltage for the intact system, but dependent on the voltage during contingencies. This can be justified by saying that the voltage deviations encountered during a standard OPF and modeled OPF are small compared to those that can be encountered during contingencies. Since the total power demand for the intact system is not changed, fuel cost comparisons between this case and a standard SCOPF seem more reasonable. We can also incorporate load models for the intact system as well as during contingencies, while minimizing the fuel cost. However, we then encounter the problem discussed in the previous section regarding net earnings. Another approach is to incorporate load models for the intact case as well as during contingencies, while minimizing the total fuel cost minus the total revenue.

Inaccuracies of Standard OPF Solutions

It was stated earlier that the standard OPF (or standard security constrained OPF) solution can give results not compatible with practical observations (i.e., using the control variable values from these solutions) when a fair number of loads are fed by fixed tap transformers. The discrepancies between the simulated and observed results will be due to discrepancies between the voltage at a bus feeding a load through a fixed tap transformer, and the voltage at which the specified power demand for that load occurs. The observed results can be simulated approximately by performing a power flow incorporating load models. The effects of load modeling in power flow studies have been treated in a few cases (Dias and El-Hawary, 1990; El-Hawary and Dias, Jan. 1987; El-Hawary and Dias, 1987; El-Hawary and Dias, July 1987). In all these studies, the specified power demand of the modeled loads was assumed to occur at a bus voltage of 1.0 per unit. The simulated modeled power flow solution will be same as the practically observed version only when exact model parameters are utilized.

SCOPF Including Load Modeling

Security constrained optimal power flow (abbreviated SCOPF) takes into account outages of certain transmission lines or equipment (Alsac and Stott, May/June 1974; Schnyder and Glavitsch, 1987). Due to the computational complexity of the problem, more work has been devoted to obtaining faster solutions requiring less storage, and practically no attention has been paid to incorporating load models in the formulations. A SCOPF solution is secure for all credible contingencies or can be made secure by corrective means. In a secure system (level 1), all load is supplied, operating limits are enforced, and no limit violations occur in a contingency. Security level 2 is one where all load is supplied, operating limits are satisfied, and any violations caused by a contingency can be corrected by control action without loss of load. Level 1 security is considered in Dias and El-Hawary (Feb. 1991).

Studies of the effects of load voltage dependence in PF and OPF (Dias and El-Hawary, Sept. 1989) concluded that for PF incorporating load models, the standard solution gives more conservative results with respect to voltages in most cases. However, exceptions have been observed in one test system. Fuel costs much lower than those associated with the standard OPF are obtained by incorporating load models with all voltages free to vary within bounds. This is due to the decrease in the power demand by the reduction of the voltages at buses whose loads are modeled. When quite a few loads are modeled, the minimum fuel costs may be much lower than the corresponding standard OPF fuel cost with a significant decrease in power demand.

A similar effect can be expected when load models are incorporated in security constrained OPF studies. The decrease in the power demand when load models are incorporated in OPF studies may not be desirable under normal operating conditions. This problem can be avoided in a security constrained OPF by incorporating load models during contingencies only. This not only gives results that are more comparable with standard OPF results, but may also give lower fuel costs without lowering the power demand of the intact system. The modeled loads are assumed to be fed by fixed tap transformers and are modeled using an exponential type of load model.

In Dias and El-Harawy (1990), some selected buses were modeled using an exponential type of load model in three cases. In the first, the specified load at modeled buses is obtained with unity voltage. In the second case, the transformer taps have been adjusted to give all industrial-type consumers 1.0 per unit at the low-voltage panel when the high-side voltage corresponds to the standard OPF solution. In the third case, the specified power demand is assumed to take place when the high-side voltages correspond to the intact case of the standard security constrained OPF solution. It is concluded that a decrease in fuel cost can be obtained in some instances when load models are incorporated in security constrained OPF studies during contingencies only. In situations where a decrease in fuel cost is obtained in this manner, the magnitude of decrease depends on the total percentage of load fed by fixed tap transformers and the sensitivity of these loads to modeling. The tap settings of these fixed tap transformers influence the results as well. An increase in fuel cost can also occur in some isolated cases. However, in either case, given accurate load models, optimal power flow solutions that are more accurate than the conventional OPF solutions can be obtained. An alternate approach for normal OPF as well as security constrained OPF is also suggested.

Influence of Fixed Tap Transformer Fed Loads

A standard OPF assumes that all loads are independent of other system variables. This implies that all loads are fed by ULTC transformers that hold the load-side voltage to within a very narrow bandwidth sufficient to justify the assumption of constant loads. However, when some loads are fed by fixed tap transformers, this assumption can result in discrepancies between the standard OPF solution and its observed version. In systems where the average voltage of the system is reasonably above 1.0 per unit (specifically where the loads fed by fixed tap transformers have voltages greater than the voltage at which the specified power demand occurs), the practically observed version of the standard OPF solution will have a higher total power demand, and hence a higher fuel cost, and total revenue, and net revenue. Conversely, where such voltages are lower than the voltage at which the specified power demand occurs, the total power demand, fuel cost, total and net revenues will be lower than expected. For the former case, the system voltages will usually be slightly less than expected, while for the latter case they will usually be slightly higher than expected.

The changes in the power demand at some buses (in the observed version) will alter the power flows on the transmission lines, and this can cause some lines to deliver more power than expected. When this occurs on transmission lines that have power flows near their upper limit, the observed power flows may be above the respective upper limit, causing a security violation. Where the specified power demand occurs at the bus voltages obtained by a standard OPF solution, the observed version of the standard OPF solution will be itself, and there will ideally be no security violations in the observed version.

Most of the above conclusions apply to security constrained OPF as well (Dias and El-Hawary, Nov. 1991). However, since a security constrained OPF solution will in general have higher voltages than its normal counterpart (in order to avoid low voltage limit violations during contingencies), the increase in power demand, and total and net revenues will be more significant while the decrease in the above quantities will be less significant. Also, the security violations due to line flows will now be experienced mainly during contingencies, as most line flows will now usually be below their upper limits for the intact case. For security constrained OPF solutions that incorporate load models only during contingencies, the simulated and observed results will mainly differ in the intact case. Also, with loads modeled during contingencies, the average voltage is lower than for the standard security constrained OPF solution and hence there will be more cases with a decrease in the power demand, fuel cost, and total and net revenues in the observed version of the results than for its standard counterpart.

Operational Requirements for Online Implementation

The most demanding requirements on OPF technology are imposed by online implementation. It was argued that OPF, as expressed in terms of smooth nonlinear programming formulations, produces results that are far too approximate descriptions of real-life conditions to lead to successful online implementations. Many OPF formulations do not have the capability to incorporate all operational considerations into the solutions. Moreover, some operating practices are occasionally incompatible with such OPF formulations. Consequently, many proposed "theoretical optimal solutions" are of little value to the operators who are almost constantly presented with simultaneous events that are outside the scope of OPF definition. These limitations, if properly addressed, do not have to prevent OPF programs from being used in practice, especially when the operational optimal solution may also not be known. Papalexopoulos (1996) offers some of the requirements that need to be met so that OPF applications are useful to, and usable by, the dispatchers in online applications.

Speed Requirements

Fast OPF programs designed for online application are needed because under normal conditions, the state of the power system changes continuously and can change abruptly during emergency conditions. The changes involve the evolution of bus active and reactive power generation and loads with time, control variables moving to and off their limits as time changes, and topology changes due to switching operations and other planned or forced outages. The need for fast OPF solutions is especially true when an excessive amount of calculations due to modeling of contingency constraints or repeated OPF runs is involved.

In general, an online OPF calculation should have been completed before the state of the power system has changed to another state that is appreciably different from the earlier state. Determining the optimal execution frequency to maximize the benefits of the computations depends on the specific situation and is limited by finite computing resources. It may be preferable to develop incrementally correct and flexible algorithms to offer fast and more frequent scheduling. This leads us to conclude that conventional formulations and algorithms characterized with quadratic convergence that give very accurate and "mathematically optimal" solutions, but neglect operational realities are not appropriate for online implementation. Fast and frequent scheduling requires "hot start" OPF capabilities developed to take advantage of the optimal status of previously optimized operating points. The hot start option is significant when the rate of change of system state is small and previously optimized points are still "relevant" to the current operating conditions.

Robustness of OPF Solutions with Respect to Initial Guess Point

An OPF program needs to produce consistent solutions and thus must not be sensitive to the selected initial guess used. In addition, changes in the OPF solutions between operating states need to be consistent with the changes in the power system operating constraints. The OPF solutions will never be exactly the same when starting from different initial guess points because the solution process is iterative. Any differences should be within the tolerances specified by the convergence criterion, and of a magnitude that would be considered insignificant to the operator. First-order OPF solution methods were not well received because noticeably different solutions could be obtained when an OPF algorithm was initialized from different initial guess points, with only one (or even none) of the solutions actually constituting a local optimum. Theoretically, if the objective function and the feasible region can be shown to be convex, then the optimal solution will be unique (Gill et al., 1981). Unfortunately, the complexity of the nonlinear equations and inequality constraints involved in OPF problems make it untenable to rigorously prove convexity. If multiple local minima actually exist, then additional computational or heuristic methods must be used to resolve the issue.

A normally feasible OPF solution space may become nonconvex (thus leading to multiple OPF solutions) due to two considerations. The first is due to use of discontinuous techniques to model specific operating practices and preferences, and the second is due to modeling of local controls. The conventional power flow problem with local control capability, whose implicit objective is feasible with respect to a limited set of inequalities, does not have a unique solution. Nevertheless, solutions of the same problem from different starting conditions usually match quite closely. Occasionally, different initial guess solutions can lead to different solutions. This takes place when two or more feasible voltage levels can satisfy nonlinear loads. OPF applications, however, should be able to overcome this type of ambiguity.

Discrete Modeling

Discrete control is widely used in the electric network. For example, transformers are used for voltage control, shunt capacitors and reactors are switched on or off to correct voltage profiles and to reduce active power transmission losses, and phase shifters are used to regulate the MW flows of transmission lines. An efficient and effective OPF discretization procedure is needed to assist the operators in utilizing discrete controls in a realistic and optimal or near-optimal manner. Discrete elements to be included in the OPF formulation are branch switching; prohibited zones of generator cost curves; and priority sequence levels for unfeasibility handling. OPF algorithms designed for online applications should be able to appropriately handle the discrete aspects of the problem.

Using both discrete and continuous controls converts the OPF into a mixed discrete-continuous optimization problem. A possible accurate solution using a method such as mixed-integer nonlinear programming would be orders of magnitude slower than ordinary nonlinear programming methods (Gill et al., 1981). Linear programming-based OPF algorithms allow substantial recognition of discrete controls by setting the cost curve segment break points at discrete control steps. However, most methods that solve for a nonseparable objective function by nonlinear programming methods do not properly model discrete controls.

Current OPF algorithms treat all controls as continuous variables during the initial solution process. Once the continuous solution is obtained, each discrete variable is moved to the nearest discrete setting. This produces acceptable solutions, assuming that the step sizes for the discrete controls are sufficiently small, which is usually the case for transformer taps and phase shifter angles (Papalexopoulos et al., 1996). Approximate solutions that can produce near-optimal results appear to be a reasonable alternative to rigorous solution methods. One such scheme (Liu et al., 1991) uses penalty functions for discrete controls. The object is to penalize the continuous approximations of discrete control variables for movements away from their discrete steps. This scheme is well suited for Newton-based OPF algorithms. The scheme consists of a set of rules to determine the timing of introduction and criteria of updating the penalties in the optimization process. This heuristic algorithm is of limited scope. Substantially more work is needed to effectively resolve all problems associated with the discrete nature of controls and other discrete elements of the OPF problem.

Detecting and Handling Infeasibility

As the requirements for satisfactory system operation increase, the region of feasible solutions that satisfy all constraints simultaneously may become too small. In this case, there is a need to establish criteria to prioritize the constraints. For OPF applications, this means that when a feasible solution cannot be found, it is still very important for the algorithm to suggest the "best optimal" engineering solution in some sense, even though it is infeasible. This is even more critical for OPF applications that incorporate contingency constraints.

There are several approaches to deal with this problem. In one approach, all power flow equations are satisfied and only the soft constraints that truly cause the bottlenecks are allowed to be violated using a least squares approximation process. An LP approach introduces a weighted slack variable for each binding constraint. If a constraint can be enforced, the slack variable will be reduced to zero and the constraint will be satisfied. The constraints causing infeasibility will have non-zero slack variables whose magnitudes are proportional to the amounts they need to be relaxed to achieve feasibility. Usually, all binding constraints of a particular type are modeled as if they have identical infeasibility characteristics. That is, all slack variables corresponding to these binding constraints share the same cost curve, and their sensitivities are scaled by a weighting factor associated with the type of the corresponding constraint. Using Newton's method, if the OPF does not converge in the first specified set of iterations, the constraint weighting factors, corresponding to the penalty functions associated with the load bus voltage limits and the branch flow limits, will be reduced successively until a solution is reached. This normally results in all constraints being met except for those load bus voltage and branch flow limits that contribute to infeasibility. Special care should be taken in selecting the proper weighting factors to avoid numerical problems and produce acceptable solutions.

Another approach develops hierarchical rules that operate on the controls and constraints of the OPF problem. The rules introduce discontinuous changes in the original OPF formulation. These changes include using a different set of control/constraint limits, expansion of the control set by class or individually, branch switching, load shedding, etc. They are usually implemented in a predefined priority sequence to be consistent with utility practices. The decision as to when to proceed to the next priority level of modifications to achieve feasibility is critical, especially when it involves radial overloads, normally overloaded constraints and constraints known to have "soft" limits. The selection of a final optimal solution among all the others in the set is achieved with the implementation of a "preference index." An application of the preference index approach that minimizes postcontingency line overloads due to generator outages is given in (Yokoyama et al., 1988).

Consistency of OPF Solutions with Other Online Functions

Online OPF is implemented in either study or closed-loop mode. In study mode, the OPF solutions are presented as recommendations to the operator. In closed-loop mode, control actions are implemented in the system via the SCADA system of the EMS (*IEEE Trans.*, June, 1983). In closed-loop mode, OPF is triggered by a number of events, including an operator request, the execution of the real-time sequence and security analysis, structural change, large load change, etc. A major concern for an OPF in closed-loop mode is the design of its interface with the other online functions, which are executed at different frequencies. Some of these functions are unit commitment, economic dispatch (ED), real-time sequence, security analysis, automatic generation control (AGC), etc. To reduce the discrepancy between ideal and realistic OPF solutions, emphasis should be placed on establishing consistency between these functions and static optimal solutions produced by OPF. This requires proper interfacing and integration of OPF with these functions. The integration design should be flexible enough to allow OPF formulation modifications consistent with the ever dynamic and sometimes ill-defined security problem definition.

Ineffective "Optimal" Rescheduling

Production-grade OPF algorithms use all available control actions to obtain an optimal solution, but for many applications it is not practical to execute more than a limited number of control actions. The OPF problem then becomes one of selecting the best set of actions of a limited size out of a much larger set of possible actions. The problem was identified but no concrete remedies were offered. It is not possible to select the best and most effective set of a given size from existing OPF solutions that use all controls to solve each problem. The control actions cannot be ranked and the effectiveness of an action is not related to its magnitude. Each control facility participates in both minimization of the objective function and enforcement of the constraints. Separation of the two effects for evaluation purposes is not feasible. The problem is difficult to define analytically and existing conventional technologies are not adequate. It is important to note that emerging computational intelligence tools such as fuzzy reasoning and neural networks may offer some resolution. The problem of ineffective rescheduling is related to but is not identical to the "minimum number of controls" objective. It is also closely linked to the problem of discrete control variables, since methods that recognize the discrete nature of some control facilities tend to decrease the number of control actions by keeping inefficient discrete controls at their initial settings.

OPF-Based Transmission Service Pricing

OPF programs are capable of computing marginal costs. Information about the optimal states with respect to changes, such as load variations, operating limit changes, or constraint parameter changes, can be used in many practical applications. Specifically, the sensitivities of the production cost of generation with respect to changes in the bus active power injections are called Bus Incremental Costs (BICs). BICs can be used as nodal prices for pricing transmission services, as they reflect the transmission loss and the congestion components for transferring power from one point to another. In a lossless network with no binding constraints, all BICs should be equal. However, when an operating limit is reached, the congestion component takes effect and all BICs in the network can be different. This means that nodal price differences across uncongested lines can be much larger than marginal losses. Extensive experience has shown that it is possible for power to flow from a bus with higher nodal price to a bus with lower nodal price, resulting in negative transmission charges. Failure to properly account for this effect can lead to unacceptable incentives for transmission users. The same applies in the case of transmission reinforcements to mitigate congestion. If as a result of the upgrades, the incremental transmission rights (positive or negative) are not accounted for properly, similar distortions are possible.

Conclusions

A review of recent developments in optimal economic operation of electric power systems with emphasis on the optimal power flow formulation was given. We dealt with conventional formulations of economic dispatch, conventional optimal power flow, and accounting for the dependence of the power demand on voltages in the system. Challenges to OPF formulations and solution methodologies for online application were also outlined.

References

Abadie, J. and Carpentier, J., Generalization of the Wolfe reduced gradient method to the case of nonlinear constraints, *Optimization,* R. Fletcher, Ed., Academic Press, New York, 1969, 37–47.

Alsac, O. and Stott, B., Optimal load flow with steady-state security, *IEEE Trans. Power App. Syst.,* PAS-93, 745–751, May/June 1974.

Fiacco, A.V. and McCormick, G.P., Computational algorithm for the sequential unconstrained minimization technique for nonlinear programming, *Management Sci.,* 10, 601–617, 1964.

Benthall, T.P., Automatic load scheduling in a multiarea power system, *Proc. IEE,* 115, 592–596, April 1968.

Burchett, R.C., Happ, H.H., Vierath, D.R., and Wirgau, K.A., Developments in optimal power flow, *IEEE Trans. Power App. and Syst.,* PAS-101, 406–414, February 1982.

Burchett, R.C., Happ, H.H., and Vierath, D.R., Quadratically convergent optimal power flow, *IEEE Trans. Power App. and Syst.,* PAS-103, 3267–3275, November 1984.

Carpentier, J., Contribution a l'etude du dispatching economique, *Bull. Soc. Francaise Electriciens,* 8, 431–447, 1962.

Carpentier, J., Differential Injections Method: A general method for secure and optimal load flows, in *IEEE PICA Conference Proceedings* Minneapolis, MN, June 1973, 255–262.

Concordia, C. and Ihara, S., Load representation in power system stability studies, *IEEE Trans. Power App. and Syst.,* PAS-101, 1982.

Contaxis, G.C., Delkis, C., and Korres, G., Decoupled optimal power flow using linear or quadratic programming, *IEEE Trans. Power Syst.,* PWRS-1, 1–7, May 1986.

Dias, L.G. and El-Hawary, M.E., A comparison of load models and their effects on the convergence of Newton power flows, *Inter. J. of Electr. Power and Energy Syst.,* 12, 3–8, 1990.

Dias, L.G. and El-Hawary, M.E., Effects of active and reactive modelling in optimal load flow studies, *IEE Proceedings,* 136, Part C, 259–263, September 1989.

Dias, L.G. and El-Hawary, M.E., Effects of load modeling in security constrained OPF studies, *IEEE Trans. on Power Systems,* 6, 1, 87–93, February 1991.

Dias, L.G. and El-Hawary, M.E., Security constrained OPF: influence of fixed tap transformer loads, *IEEE Trans. on Power Systems,* 6, 1366–1372, November 1991.

Dommel, H.W. and Tinney, W.F.,Optimal power flow solutions, *IEEE Trans. Power App. and Syst.,* PAS-87, 1866–1876, October 1968.

El-Din, H.M.Z., Burns, S.D., and Graham, C.E., Voltage/var control with limited control actions, in *Proceedings of the Canadian Electrical Association Spring Meeting,* Toronto, Ontario, March 1989.

El-Hawary, M.E., Power system load modeling and incorporation in load flow solutions, in *Proceedings of the Third Large Systems Symposium,* University of Calgary, June 1982.

El-Hawary, M.E. and Dias, L.G., Bus sensitivity to model parameters in load-flow studies, *IEE Proceedings,* 134, Part C, 302–305 July 1987.

El-Hawary, M.E. and Dias, L.G., Incorporation of load models in load flow studies: form of model effects, *IEE Proceedings,* 134, Part C, 27–30 January 1987.

El-Hawary, M.E. and Dias, L.G., Selection of buses for detailed modeling in load flow studies, *Electric Machines and Power Syst.,* 12, 83–92, 1987.

El-Kady, M.A., Bell, B.D., Carvalho, V.F., Burchett, R.C., Happ, H.H., and Vierath, D.R., Assessment of real-time optimal voltage control, *IEEE Trans. Power Systems,* PWRS-1, 98–107, May 1986.

Fletcher, R. and Powell, M.J.D., A rapidly convergent descent method for minimization, *Computer Journal,* 6, 163–168, 1963.
Gill, P.E., Murray, W., and Wright, M.H., *Practical Optimization*, Academic Press, New York, 1981.
Glavitsch, H. and Spoerry, M., Quadratic loss formula for reactive dispatch, *IEEE Trans. Power App. and Syst.,* PAS-102, 3850–3858, December 1983.
Happ, H.H. and Vierath, D.R., The OPF for operations planning and for use on line, in *Proceedings of the Second International Conference on Power Systems Monitoring and Control,* July 1986, 290–295.
IEEE Committee Report, System load dynamic simulation, effects and determination of load constants, *IEEE Trans. Power App. and Syst.,* PAS-92, 600–609, 1973.
IEEE Current Operating Problems Working Group Report, On-line load flows from a system operator's viewpoint, *IEEE Trans. on Power Appar. Syst.*, PWRS-102, 1818–1822, June 1983.
IEEE Working Group Report, The effect of frequency and voltage on power system loads, Paper 31 CP 66-64, IEEE Winter Meeting, New York, 1966.
Iliceto, F., Ceyhan, A., and Ruckstuhl, G., Behavior of loads during voltage dips encountered in stability studies, *IEEE Trans. Power App. and Syst.*, PAS-91, 2470–2479, 1972.
Karmarkar, N., New polynomial-time algorithm for linear programming, *Combinatorica,* 4, 1984, 00.373-395.
Kuhn, H.W. and Tucker, A.W., Nonlinear programming, in *Proc. of Second Berkeley Symposium on Mathematical Statistics and Probability,* University of California, Berkeley, 1951, 481–492.
Lindqvist, A., Bubenko, J.A., and Sjelvgren, D., A generalized reduced gradient methodology for optimal reactive power flows, in *Proceedings of the* 8*th PSCC,* Helsinki, Finland, 1984.
Liu, E., Papalexopoulos, A.D., and Tinney, W.F., Discrete shunt controls in a Newton optimal power flow, *IEEE Winter Power Meeting 1991,* Paper 91WM 041-4 PWRS.
Lootsma, F.A., Logarithmic programming: a method of solving nonlinear programming problems, *Philips Res. Repts.,* 22, 329–344, 1967.
Maria, G.A. and Findlay, J.A., A Newton optimal power flow program for Ontario hydro EMS, *IEEE Trans. Power Syst.,* PWRS-2, 576–584, Aug. 1987.
Merlin, A., On optimal generation planning in large transmission systems (The Maya Problem), in *Proc. of* 4*th PSCC,* Grenoble, paper 2.1/6, September 1972.
Momoh, J.A., Application of Quadratic Interior Point Algorithm to Optimal Power Flow, EPRI Final report RP 2473-36 II, March 1992.
Papalexopoulos, A. D., Challenges to an on-line OPF implementation, in *IEEE Tutorial Course Optimal Power Flow: Solution Techniques, Requirements, and Challenges,* Publication 96 TP 111-0, IEEE Power Engineering Society, 1996.
Papalexopoulos, A.D., Imparato, C.F., and Wu, F.F., Large-scale optimal power flow: effects of initialization, decoupling and discretization, *IEEE Trans. on Power Appar. Syst.,* PWRS-4, 748–759, May 1989.
Peschon, J., Bree, D.W., and Hajdu, L.P., Optimal solutions involving system security, in *Proc. of the* 7*th PICA Conf.,* Boston, MA, 1971, 210–218.
Polak, E. and Ribiere, G., Note sur la convergence de methods de directions conjugees, *Rev. Fr. Inform. Rech. Operation,* 16-R1, 35–43, 1969.
Sasson, A.M., Combined use of the Powell and Fletcher-Powell nonlinear programming methods for optimal load flows, *IEEE Trans. Power App. and Syst.,* PAS-88, 1530–1537, October 1969.
Sasson, A.M., Viloria, F., and Aboytes, F., Optimal load flow solution using the Hessian matrix, *IEEE Trans. Power App. and Syst.,* PAS-92, 31–41, Jan/Feb 1973.
Schnyder, G. and Glavitsch, H., Integrated security control using an optimal power flow and switching concepts, in *IEEE PICA Conf. Proc.,* Montreal, Canada, May 18–22, 1987, 429–436.
Shen, C.M. and Laughton, M.A., Power system load scheduling with security constraints using dual linear programming, *IEE Proceedings,* 117, 2117–2127, Nov. 1970.
Stott, B., Alsac, O., and Monticelli, A. J., Security analysis and optimization, *Proc. IEEE,* 75, 1623–1644, Dec. 1987.

Stott, B. and Hobson, E., Power system security control calculations using linear programming. Parts 1 and 2, *IEEE Trans. Power App. and Syst.,* PAS-97, 1713–1731, Sept./Oct. 1978.
Sun, D.I., Ashley, B., Brewer, B., Hughes, A., and Tinney, W.F., Optimal power flow by Newton approach, *IEEE Trans. Power App. and Syst.,* PAS-103, 2864–2880, Oct. 1984.
Vaahedi, E. and El-Din, H.M.Z., Considerations in applying optimal power flow to power systems operation, *IEEE Trans. Power Syst.,* PWRS-4, 694–703, May 1989.
Vaahedi, E., El-Kady, M., Libaque-Esaine, J.A., and Carvalho, V.F., Load models for large scale stability studies from end user consumption, *IEEE Trans. Power Syst.,* PWRS-2, 864–871, 1987.
Vargas, L.S., Quintana, V.H., and Vannelli, A., A tutorial description of an interior point method and its applications to security-constrained economic dispatch, *IEEE Trans. on Power Systems,* 8, 1315–1323, 1993.
Wells, D.W., Method for economic secure loading of a power system, *IEE Proceedings,* 115, 1190–1194, Aug. 1968.
Wolfe, P., Methods of nonlinear programming, *Nonlinear Programming,* J. Abadie, Ed., North Holland, Amsterdam, 1967, 97–131.
Yan, X. and Quintana, V. H., Improving an interior-point-based OPF by dynamic adjustments of step sizes and tolerances, *IEEE Trans. on Power Systems,* 14, 709–717, 1999.
Yokoyama, R., Bae, S.H., Morita, T., and Sasaki, H., Multiobjective optimal generation dispatch based on. probability security criteria, *IEEE Trans. on Power Appar. Syst.,* PWRS-3, 317–324, 1988.
Yu, D.C., Fagan, J.E., Foote, B., and Aly, A.A., An optimal load flow study by the generalized reduced gradient approach, *Electric Power Systems Research,* 10, 47–53, 1986.
Zangwill, W.I., Non-linear programming via penalty functions, *Management Sci.,* 13, 344–358, 1967.

12.5 Security Analysis

Nouredine Hadjsaid

The power system as a single entity is considered the most complex system ever built. It consists of various equipment with different levels of sophistication, complex and nonlinear loads, various generations with a wide variety of dynamic responses, a large-scale protection system, a wide-area communication network, and numerous control devices and control centers. This equipment is connected with a large network (transformers, transmission lines) where a significant amount of energy transfer often occurs. This system, in addition to the assurance of good operation of its various equipment, is characterized by an important and simple rule: electricity should be delivered to where it is required in due time and with appropriate features such as frequency and voltage quality. Environmental constraints, the high cost of transmission investments and low/long capital recovery, and the willing of utilities to optimize their network for more cost effectiveness makes it very difficult to expand or oversize power systems. These constraints have pushed power systems to be operated close to their technical limits, thus reducing security margins.

On the other hand, power systems are continuously subjected to random and various disturbances that may, under certain circumstances, lead to inappropriate or unacceptable operation and system conditions. These effects may include cascading outages, system separation, widespread outages, violation of emergency limits of line current, bus voltages, system frequency, and loss of synchronism (Debs and Benson, 1975). Furthermore, despite advanced supervisory control and data acquisition systems that help the operator to control system equipment (circuit breakers, on-line tap changers, compensation and control devices, etc.), changes can occur so fast that the operator may not have enough time to ensure system security. Hence, it is important for the operator not only to maintain the state of the system within acceptable and secure operating conditions but also to integrate preventive functions. These functions should allow him enough time to optimize his system (reduction of the probability of occurrence of abnormal or critical situations) and to ensure recovery of a safe and secure situation.

Even though for small-scale systems the operator may eventually, on the basis of his experience, prevent the consequences of most common outages and determine the appropriate means to restore a secure state, this is almost impossible for large systems. It is therefore essential for operators to have at their disposal, efficient tools capable of handling a systematic security analysis. This can be achieved through the diagnosis of all contingencies that may have serious consequences. This is the concern of **security analysis**.

The term contingency is related to the possibility of losing any component of the system, whether it is a transmission line, a transformer, or a generator. Another important event that may be included in this definition concerns busbar faults (bus split). This kind of event is, however, considered rare but with (serious) dangerous consequences. Most power systems are characterized by the well-known N – 1 security rules where N is the total number of system components. This rule is the basic requirement for the planning stage where the system should be designed in order to withstand (or to remain in a normal state) any single contingency. Some systems also consider the possibility of N – 2/k (k is the number of contingencies), but mostly for selected and specific cases.

Definition

The term security as defined by NERC (1997) is the ability of the electric systems to withstand sudden disturbance such as electric short-circuits or unanticipated loss of system elements. (See Appendix A).

Security analysis is usually handled for two time frames: static and dynamic. For the static analysis, only a "fixed picture" or a snapshot of the network is considered. The system is supposed to have passed the transient period successfully or be dynamically stable. Therefore, the monitored variables are line flows and bus voltages. Hence, all voltages should be within a predefined secure range, usually around ±5% of nominal voltage (for some systems, such as distribution networks, the range may be wider). In fact, if bus voltages drop below a certain level, there will be a risk of voltage collapse in addition to high losses. On the other hand, if bus voltages are too high compared to nominal values, there will be equipment degradation or damage. Furthermore, overload of transmission lines may be followed by unpredictable line tripping that accelerates the degradation of the voltage profile.

Line flows are related to circuit overload (lines and transformers) and should keep below a maximum limit, usually settled according to line thermal limits. The dynamic security is related to loss of synchronism (transient stability) and oscillatory swings or dynamic instability. In that case the evolution of essential variables are monitored based upon a required time frame (transient period).

Normally, system security is analyzed differently whether it is considered for planning studies or for monitoring and operational purposes. The difference comes from the type of action that should be initiated in case of expected harmful contingencies. However, for both stages, all variables should remain within the bounded domain defining or determining system normal state (Fink, 1978).

Time Frames for Security-Related Decision (McCaulley et al., 1999)

There are generally three different time frames for security-related decisions. In operations, the decision-maker is the operator, who must continuously monitor and operate his system economically in such a way that the normal state is appropriately preserved (maintained). For this purpose, he has specific tools for diagnosing his system and operating rules that allow the required decisions to be made in due time. In operational planning, the operating rules are developed recognizing that the bases for the decision are reliability/security criteria specifying minimum operating requirements, which define acceptable performance for the credible contingencies. In facility planning, the planner must determine the best way to reinforce the transmission system, based on reliability/security criteria for system design, which generally adhere to the same disturbance-performance criteria specified by minimum operating requirements.

One may think that since these systems are designed to operate "normally" or in "a secure state" for a given security rule (N – k), there is nothing to worry about during operations. The problem is that, during the planning stage and for a set of given economical constraints, a number of assumptions are made for operating conditions that concern topology, generation, and consumption. Since there may be several years between the planning stage and the operations, the uncertainties in the system's security may be very significant. Therefore, security analysis is supplemented by operational planning and operations studies.

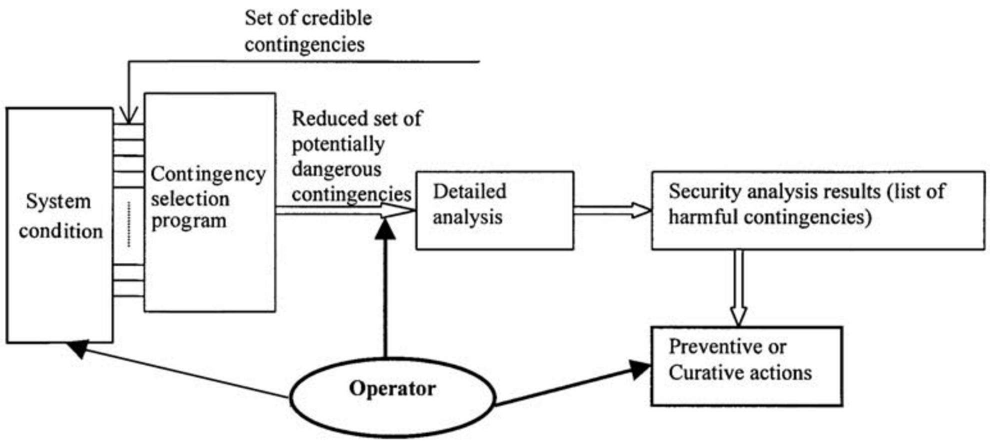


FIGURE 12.21 Contingency analysis procedure.

The decision following any security analysis can be placed in one of two categories: preventive or corrective actions. For corrective actions, once a contingency or an event is determined as potentially dangerous, the operator should be confident that in case of that event, he will be able to correct the system by means of appropriate actions on system conditions (generation, load, topology) in order to keep the system in a normal state and even away from the insecure region. The operator should also prepare a set of preventive actions that may correct the effect of the expected dangerous event.

In operations, the main constraint is the time required for the analysis of the system's state and for the required decision to be made following the security analysis results. The security analysis program should be able to handle all possible contingencies, usually on the N – 1 basis or on specific N – 2. For most utilities, the total time window considered for this task is between 10 min and 30 min. Actually for this time window, the system's state is considered as constant or quasi-constant allowing the analysis to be valid within this time frame. This means that changes in generation or in consumption are considered as negligible.

For large systems, this time frame is too short even with very powerful computers. Since it is known that only a small number of contingencies may really cause system violations, it has been realized that it is not necessary to perform a detailed analysis on all possible contingencies, which may be on the order of thousands. For this purpose, the operator may use his engineering judgment to select those contingencies that are most likely to cause system violation. This procedure has been used (and is still in use) for many years in many control centers around the world. However, as system conditions are characterized by numerous uncertainties, this approach may not be very efficient especially for large systems.

The concept of contingency selection has arisen in order to reduce the list of all possible contingencies to only the potentially harmful. The selection process should be very fast and accurate enough to identify dangerous cases (Hadjsaid, 1992). This process has existed for many years, and still is a major issue in all security studies for operations whether for static or dynamic and transient purposes.

Models

The static security analysis is mainly based on load flow equations. Usually, active/angle and reactive/voltage problems are viewed as decoupled. The active/angle subproblem is expressed as (Stott and Alsac, 1974):

$$\Delta\theta = \left[dP/d\theta\right]^{-1} \Delta P \tag{12.53}$$

where $\Delta\theta$ is a vector of angular changes with a dimension of Nb – 1 (Nb = number of buses), ΔP a vector of active injection changes (Nb – 1) and $[dP/d\theta]$ is a part of the Jacobian matrix. In the DC approach, this Jacobian is approximated by the B′ (susceptance) matrix representing the imaginary part of the Ybus matrix. This expression is used to calculate the updated angles following a loss of any system component. With appropriate numerical techniques, it is straightforward to update only necessary elements of the equation. Once the angles are calculated, the power flows of all lines can be deducted. Hence, it is possible to check for line limit violation.

Another approach that has been, and still is used in many utilities for assessing the impact of any contingency on line flows is known as shift factors. The principle used recognizes that the outage of any line will result in a redistribution of the power previously flowing through this line on all the remaining lines. This distribution is mainly affected by the topology of the network. Hence, the power flow of any line ij following an outage of line km can be expressed as (Galiana, 1984) (see Appendix B for more details):

$$P_{ij/km} = P_{ij} + \alpha_{ij/km} * P_{km} \tag{12.54}$$

where $P_{ij/km}$ is the active power flow on line ij after the outage of line km

P_{ij}, P_{km} is the active power previously flowing respectively on line ij and km (before the outage)

$\alpha_{ij/km}$ is the shift factor for line ij following the outage of line km

Equation (12.54) shows that the power flow of line ij ($P_{ij/km}$) when line km is tripped, is determined as the initial power flow on line ij (P_{ij}) before the outage of line km plus a proportion of the power flow previously flowing on line km. This proportion is defined by the terms $\alpha_{ij/km}$ * P_{km}.

The shift factors are determined in a matrix form. The important features of these factors are the simplicity of computing and their dependency on network topology. Therefore, if the topology does not change, the factors remain constant for any operating point. The main drawback of these factors is that they are determined on the basis of DC approximation and the shift factor matrix should be updated for any change in the topology. In addition, for some complex disturbances such as bus split, updating these factors becomes a complicated task.

A similar method based on reactive power shift factors has been developed. Interested readers may refer to Ilic-Spong and Phadke (1986) and Taylor and Maahs (1991) for more details.

The reactive/voltage subproblem can be viewed as (Stott and Alsac, 1974):

$$\Delta V = \left[dQ/dV\right]^{-1} \Delta Q \tag{12.55}$$

where ΔV is the vector of voltages change (Nb – Ng, Ng is the number of generators)

ΔQ is the vector of reactive power injections change (Nb – Ng, Ng is the number of generators)

[dQ/dV] is the Jacobean submatrix

In the well-known FDLF (Fast Decoupled Load Flow) model (Stott and Alsac, 1974), the Jacobian submatrix is replaced by the B″ (susceptance) matrix representing the imaginary part of the Ybus matrix with a dimension of Nb – Ng, where Ng is the number of voltage regulated (generator) buses. In addition, the vector ΔQ is replaced by $\Delta Q/V$.

Once bus voltages are updated to account for the outage, the limit violations are checked and the contingency effects on bus voltages can be assessed.

The most common framework for the contingency analysis is to use approximate models for the selection process, such as the DC model, and use the AC power flow model for the evaluation of the actual impact of the given contingency on line flows and bus voltages.

Concerning the dynamic security analysis, the framework is similar to the one in static analysis in terms of selection and evaluation. The selection process uses simplified models, such as Transient Energy Functions (TEF), and the evaluation one uses detailed assessing tools such as time domain simulations. The fact that the dynamic aspect is more related to transient/dynamic stability technique makes the process much more complicated than for the static problem. In fact, in addition to the number of contingencies to be analyzed, each analysis will require detailed stability calculations with an appropriate network and system component model such as the generator model (park, saturation, etc.), exciter (AVR: Automatic Voltage Regulator; PSS: Power System Stabilizer), governor (nuclear, thermal, hydroelectric, etc.), or loads (non-linear, constant power characteristics, etc.). In addition, integration and numerical solutions are an important aspect for these analyses.

Determinist vs. Probabilistic

The basic requirement for security analysis is to assess the impact of any possible contingency on system performance. For the purpose of setting planning and operating rules that will enable the system to be operated in a secure manner, it is necessary to consider all credible contingencies, different network configurations, and different operating points for given performance criteria. Hence, in the deterministic approach, these assessments may involve a large number of computer simulations even if there is a selection process at each stage of the analysis. The decision in that case is founded on the requirement that each outage event in a specified list, the contingency set, results in system performance that satisfies the criteria of the chosen performance evaluation (Fink and Carlsen, 1978). To handle these assessments for all possible situations by an exhaustive study is generally not reasonable. Since the resulting security rules may lead to the settlement and schedule of investment needs as well as operating rules, it is important to optimize the economical impact of security measures that have to be taken in order to be sure that there is no unnecessary or unjustified investment or operating costs. This has been the case for many years, since the emphasis was on the most severe, credible event leading to overly conservative solutions.

One way to deal with this problem is the concept of the probability of occurrence (contingencies) in the early stage of security analysis. This can be jointly used with a statistical approach (Schlumberger et al., 1999) that allows the generation of appropriate scenarios in order to fit more with the reality of the power system from the technical point of view as well as from the economical point view.

Security under Deregulation

With deregulation, the power industry has pointed out the necessity to optimize the operations of their systems leading to less investment in new facilities and pushing the system to be exploited closer to its limits. Furthermore, the open access has resulted in increased power exchanges over the interconnections. In some utilities, the number of transactions previously processed in one year is now managed in one day. These increased transactions and power exchanges have resulted in increased parallel flows leading to unpredictable loading conditions or voltage problems. A significant number of these transactions are non-firm and volatile. Hence, the security can no longer be handled on a zonal basis but rather on large interconnected systems.

Appendix A

The current NERC basic reliability requirement from NERC Policy 2- transmission (Pope, 1999) is:

Standards

1. Basic reliability requirement regarding single contingencies: All control areas shall operate so that instability, uncontrolled separation, or cascading outages will not occur as a result of the most severe single contingency.
   1.1 Multiple contingencies: Multiple outages of credible nature, as specified by regional policy, shall also be examined and, when practical, the control areas shall operate to protect against instability, uncontrolled separation, or cascading outages resulting from these multiple outages.
   1.2 Operating security limits: Define the acceptable operating boundaries
2. Return from Operating security limit violation: Following a contingency or other event that results in an operating security limit violation, the control area shall return its transmission system to within operating security limits soon as possible, but no longer than 30 minutes.

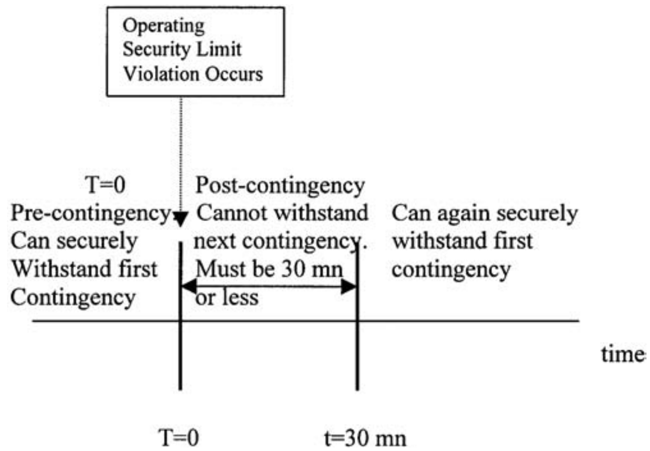


FIGURE 12.22 Current NERC basic reliability requirement. (Pope, J. W., Transmission Reliability under Restructuring, in *Proceedings of IEEE SM 1999,* Edmonton, Alberta, Canada, 162–166, July 18–22, 1999. With permission.)

Appendix B

Shift factor derivation (Galiana, 1984)
Consider a DC load flow for a base case:

$$\left[B'\right]\underline{\theta} = \underline{P}$$

where θ is the vector of phase angles for the base case
[B′] is the susceptance matrix for the base case
P is the vector of active injections for the base case

Suppose that the admittance of line jk is reduced by ΔY_{jk} and the vector ΔP is unchanged, then:

$$\left[\left[B'\right] - \Delta Y_{jk}\, \underline{e}_{\underline{jk}}\, \underline{e}_{\underline{jk}}^{\ T}\right]\underline{\theta} = \underline{P}$$

where e_{jk} is the vector (Nb – 1) containing 1 in the position j, –1 in the position k and 0 elsewhere
T is the Transpose

Now we can compute the power flow on an arbitrary line lm when line jk is outaged:

$$P_{lm/jk} = Y_{lm}\left(\theta_1 - \theta_m\right) = Y_{lm}\, \underline{e}_{\underline{lm}}^{\ T}\, \underline{\theta}$$

$$= Y_{lm}\, \underline{e}_{\underline{lm}}^{\ T}\left[\left[B'\right] - \Delta Y_{jk}\, \underline{e}_{\underline{jk}}\, \underline{e}_{\underline{jk}}^{\ T}\right]^{-1} P$$

By using the matrix inversion lemma, we can compute:

$$P_{lm/jk} = Y_{lm}\, e_{lm}^{\ T}\left[\left[B'\right] + \left(\left[B'\right]^{-1} e_{jk}\, e_{jk}^{\ T}\left[B'\right]^{-1}\right) \Big/ \left(\left(\Delta Y_{jk}\right) - 1 - e_{jk}^{\ T}\left[B'\right]^{-1} e_{jk}\right)\right] P$$

Finally:

$$P_{lm/jk} = P_{lm} + \alpha_{jk/jk} * P_{jk}$$

where

$$\alpha_{jk/jk} = Y_{lm} * \left(\Delta Y_{jk} / Y_{jk}\right) * \left(e_{lm}{}^{T} \left[B'\right]^{-1} e_{jk}\right) \Big/ \left(1 - \Delta Y_{jk}\, e_{jk}{}^{T} \left[B'\right]^{-1} e_{jk}\right)$$

References

Debs, A. S. and Benson, A. R., Security Assessment of Power Systems, in *System Engineering for Power: Status and Prospects,* Henniker, New Hampshire, 144-178, Aug. 17–22, 1975.

Fink, L. and Carlsen, K., Operating Under Stress and Strain, *IEEE Spectrum,* 15, 48–53, March, 1978.

Galiana, F. D., Bound estimates of the severity of line outages in power system contingency analysis and ranking, *IEEE Trans. on Power Appar. and Syst.,* PAS-103(9), 2612–2624, September 1984.

Hadjsaid, N., Benahmed, B., Fandino, J., Sabonnadiere, J.-Cl., and Nerin, G., Fast contingency screening for voltage-reactive considerations in security analysis, *IEEE Winter Meeting,* 1992 WM 185-9 PWRS.

Ilic-Spong, M. and Phadke, A., Redistribution of reactive power flow in contingency studies, *IEEE Trans. on Power Syst.,* PWRS-1(3), 266–275, August 1986.

McCaulley, J. D., Vittal, V., and Abi-Samra, N., An overview of risk based security assessment, in *Proceedings of IEEE SM'99,* Edmonton, Alberta, Canada, 173–178, July 18–22, 1999.

The North American Reliability Council, NERC Planning Standards, approved by NERC Board of Trustees, September, 1997.

Pope, J. W., Transmission reliability under restructuring, in *Proceedings of IEEE SM'99,* Edmonton, Alberta, Canada, 162–166, July 18–22, 1999.

Schlumberger, Y., Lebrevelec, C., and De Pasquale, M., Power system security analysis: New approaches used at EDF, in *Proceedings of IEEE SM'99,* Edmonton, Alberta, Canada, 147–151, July 18–22, 1999.

Stott, B. and Alsac, O., Fast decoupled load flow, *IEEE Trans. on Power Appar. and Syst.,* PAS-93, pp. 859-869, May/June 1974.

Taylor, D. G. and Maahs, L. J., A reactive contingency analysis algorithm using MW and MVAR distribution factors, *IEEE Trans. on Power Syst.,* 6, 349–355, February 1991.