2 Fundamental Supporting Concepts

This chapter provides an overview of the main infrastructure systems that are the focus of this book. It also describes fundamental concepts and information about network theory, reliability, and availability, and disruptive events that are also applicable to the rest of this book.

2.1 Electric Power Grid Fundamentals

Figure 2.1 displays the main components and general structure of a conventional power grid. A conventional power system has three main parts: generation, transmission, and distribution. Electrical energy is generated in power plants or power stations. These power stations are few relative to the number of loads, that is, the difference in the number of power stations to the number of loads is four or five orders of magnitude. This is a significant difference with microgrids (discussed in Chapter 6), in which the difference in the number of loads to the number of power plants is usually at most one or two orders of magnitude. Another important difference between conventional grids and microgrids is that the typical capacity of a power station is a few gigawatts provided by usually less than a dozen power generation units, each with an output of a few hundred megawatts, whereas loads are typically in the range of a few kilowatts in homes to a few hundred kilowatts for a few individual loads in industries. That is, there are several orders of magnitude difference between power generation units' capacity and individual loads' power consumption, whereas in microgrids power generation units' capacity and individual loads' power consumption are more comparable. Because of the order of magnitude difference in the number and rated power of power generation units and loads observed in conventional power grids, loads are considered at an aggregated level when planning or operating power plants. Thus, conventional power grids can be considered as a system with a mostly centralized architecture even though today there are some small power plants tied to the grid near the loads at the power distribution level of the grid – hence, they receive the name of distributed power plants. However, distributed power generation contributes a relatively very small percentage to the total electrical power generated in conventional power grids.

Power stations can be loosely separated into those using thermal energy to produce electricity and those that do not use thermal energy for electric power generation.

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

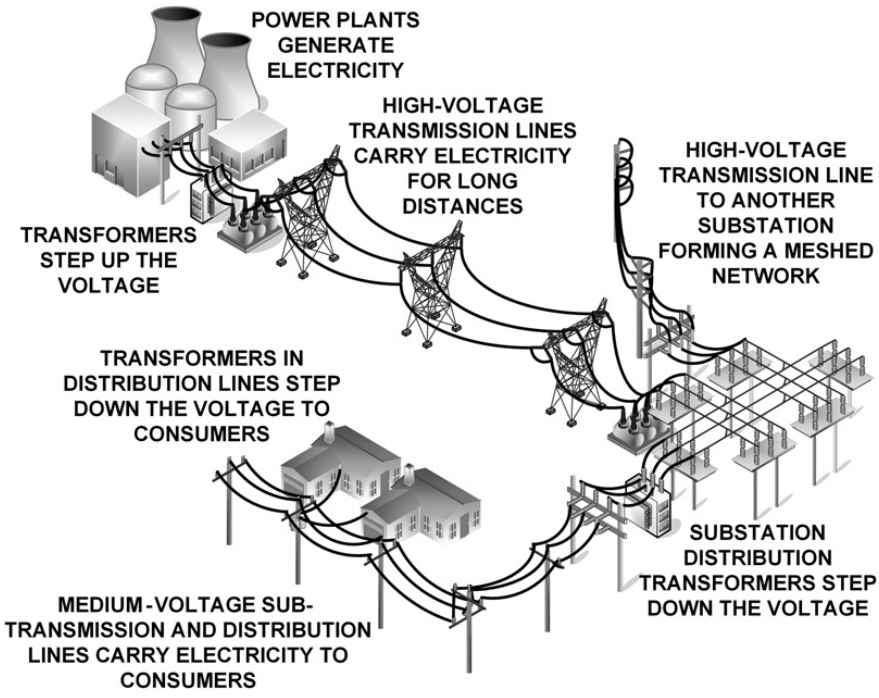

Figure 2.1 Simplified representation of the main components of a conventional power grid.

These latter power plants mainly refer to renewable energy sources. The largest type of power stations using renewable energy sources are hydroelectric power plants, which convert mainly potential energy in water, typically stored in reservoirs created with dams, to mechanical energy (work) in turbines, which, in turn, act on electric power generators that convert the mechanical energy into electrical energy. Large hydroelectric power stations have installed capacities over 1 GW of power that could reach 14 GW in the Itaipu power station and 22.5 GW in the Three Gorges Dam power stations, which are the two largest hydroelectric power plants in the world. Both these power stations produce electric power primarily with 700 MW turbines. More modest hydroelectric power stations have installed capacities of a few hundred kilowatts, such as the Cowans Ford Hydroelectric Station in Fig. 2.2 that has an installed capacity of 350 MW from its four turbines. Other forms of renewable energy power plants are solar (thermal and photovoltaics) and wind farms. Solar thermal power plants, such as that in Fig. 2.3, have installed capacities of at most a few hundred megawatts. Utility-scale photovoltaic power plants commonly have a capacity of less than 5 MW, although there are few cases of these types of power plants with a capacity of a few hundred megawatts and in very few cases just over 1 GW. Typically, wind farms have a total installed capacity of a few hundred megawatts and rarely exceed 1 GW. Wind turbines have an average individual power capacity of about 2 MW. One advantage of wind power generation systems is that wind farms could be built offshore, as exemplified in Fig. 2.4. Photovoltaic systems, however, are more suitable than wind turbines

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

Figure 2.2 Cowans Ford Hydroelectric Station. The arrows indicate the location of its four turbines.

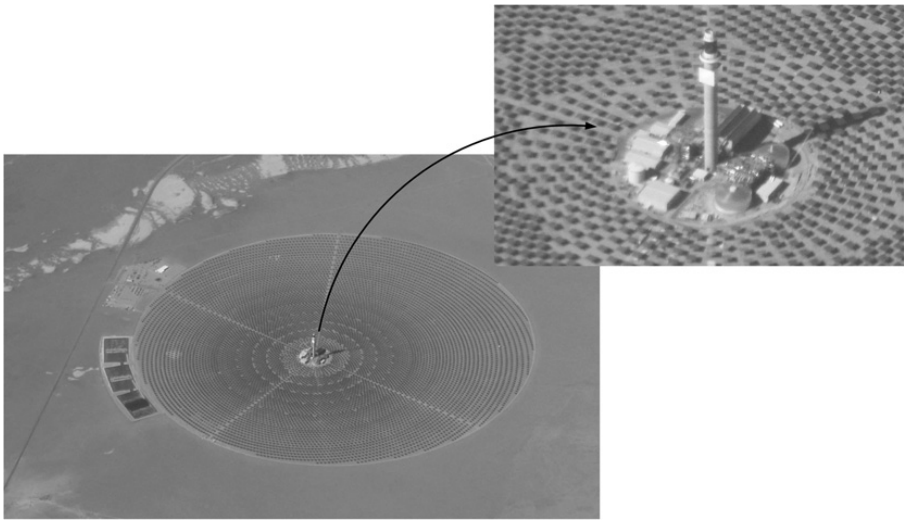

Figure 2.3 The Crescent Dunes Solar Energy Project with a capacity of 110 MW.

for their use in distributed generation applications. However, both wind and photovoltaic systems' output is either stochastic in the case of the former farms or partially stochastic in the case of the latter power generation stations. Such variable power output is not an issue with hydroelectric power stations because these power plants include energy storage through their water reservoirs. Nevertheless, on the one hand a common issue with all these renewable energy power stations is that they need to be built where their renewable energy resource is harvested, which in many cases is away

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

Figure 2.4 An offshore wind farm near Copenhagen, Denmark.

Figure 2.5 The Gloucester Marine Terminal photovoltaic power plant.

from the loads they are serving. Additionally, these renewable energy technologies, particularly photovoltaics, have a large footprint, which further complicates their installation close to the load centers. Hence, even when it is possible to find photovoltaic or wind power plants in cities, their installed capacity is limited to a few hundred kilowatts and only in a few exceptional cases, such as in Fig. 2.5, does their capacity exceed 1 MW. Thus, in most cases, practical utilization of renewable energy resources requires building transmission lines to connect these power stations to loads, as exemplified in Fig. 2.6. On the other hand, renewable energy power stations do not depend on the provision of a fuel, which, as discussed in Chapter 4, introduces dependencies that affect resilience. There are also other technologies of renewable energy power stations, such as geothermal power plants or ocean energy systems, but

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

Figure 2.6 Part of the San Gorgonio Pass wind farm (installed capacity of 615 MW) with one of the transmission lines used to connect it to the load centers.

Figure 2.7 The Donald Von Raesfeld Power Plant equipped with two 50 MW natural gas-fueled combustion turbines and one 22 MW steam turbine.

none of these other technologies have been implemented as extensively as hydroelectric, wind, or solar energy systems.

In thermal power plants, prime movers (usually a turbine) convert thermal energy to mechanical energy – namely, work – which, in turn, is converted to electrical energy in power generators in which the rotor is mounted on the same shaft as the prime movers. Thermal power stations can then be classified with respect to the prime mover technology including their fuel. In coal, fuel oil, and nuclear power plants, steam generated in boilers or heat exchangers drives the turbines. In natural gas turbines, a combustion process within the turbine creates the torque that results in the output work without the need of using steam. In combined-cycle power plants, such as the one in Fig. 2.7, the hot

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

Figure 2.8 The Linden Cogeneration Power Plant.

exhaust of the natural gas turbines is passed through a heat exchanger/boiler that produces steam to drive a second turbine also used to generate electrical power. Another example of a combined-cycle power plant is shown in Fig. 2.8. This power plant is equipped with six gas-fired electric power generation units totaling 972 MW. Units one to five are equal combined-cycle units with a combined nameplate output of 800 MW. Each of these natural gas turbines has a simple-cycle capacity of 90 MW and can also operate with butane. Their output is injected into the grid through a 345 kV transmission line. The remaining power generation turbine, unit #6, operates with natural gas as well as distillate and is being modified to operate with mixed fuel by adding up to 40 percent hydrogen. Large natural gas power stations, such as the one in Fig. 2.9, have installed capacities of a few gigawatts, but regular-size natural gas power plants, such as the one in Fig. 2.10, have installed capacities of a few hundred megawatts. Because of its high demand for fuel, the power station in Fig. 2.9 also serves as a natural gas terminal, which includes storage tanks.

As indicated, the need to have fuel delivered to power stations not using renewable energy sources introduces dependencies that may negatively affect resilience. One alternative to reduce this effect of dependencies is to use a variety of fuels, such as the case of the Linden Cogeneration Plant in Fig. 2.8 or the EF Barret Power Plant in Fig. 2.11, which, although it is fueled in all of its units primarily by natural gas, it also accepts fuel oil #6. Another example of a power plant that uses fuel oil in some of its units as a secondary fuel to natural gas is the one in Figure 2.12, which has this option for power generation units 1 to 3.

During operation under normal conditions, power stations' output is usually determined based on an optimal economic dispatch strategy that includes constraints, such

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

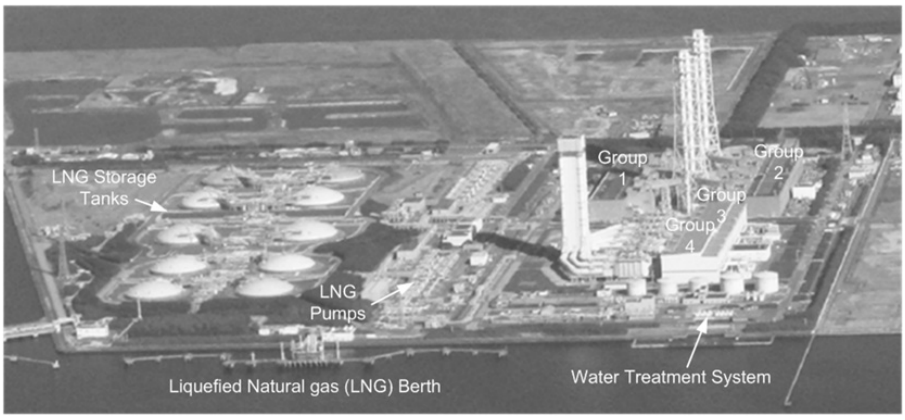

Figure 2.9 The Futtsu Power Station equipped with two groups of 1,520 MW natural gas–fueled generators and two groups of 1,000 MW natural gas–fueled generators for a total installed capacity of 5,040 MW.

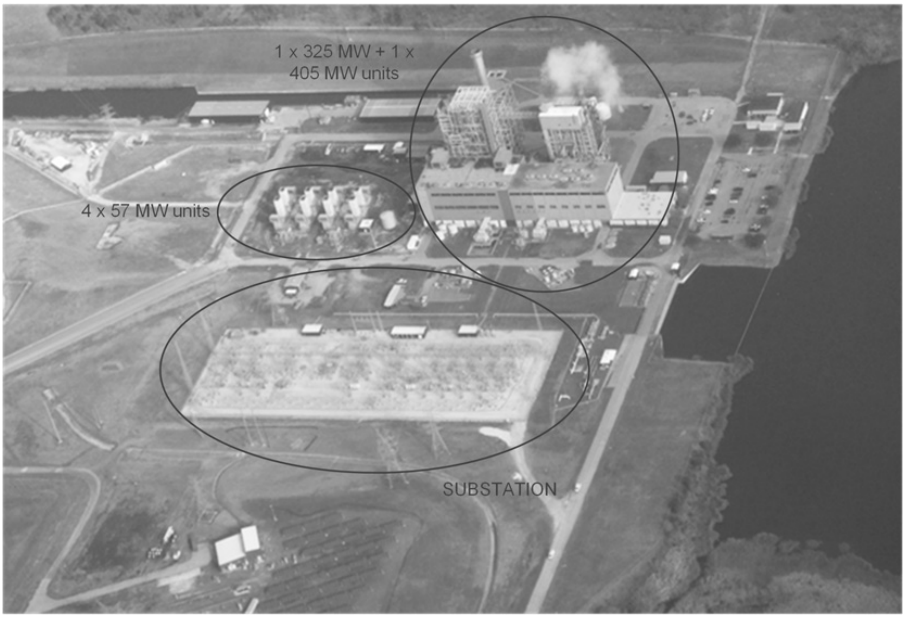

Figure 2.10 The Decker Creek power station with a total capacity in its natural gas turbines of 958 MW. The photovoltaic system at the bottom of the image has a capacity of 300 kW and has about the same footprint as the four 57 MW natural gas turbines.

as power output limits and power grid reserves' needs to ensure stable operation. In the case of the Irsching power station in Fig. 2.12, its owner requested to have it decommissioned because its operation was relatively costly. However, grid operators denied this request because such a power plant was necessary to ensure adequate grid operation. Thus, its generators were placed in standby reserve. This case exemplifies

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

Figure 2.11 E. F. Barrett Power Plant, which is comprised of two 195 MW dual-fuel steam units and eleven simple-cycle, dual-fuel combustion turbines, which are divided between fifteen 15 MW and four 40 MW combustion turbines.

Figure 2.12 Irsching Power Station, which operates five power generation units: Unit #1 of 150 MW, Unit #2 of 330 MW, Unit #3 of 440 MW, Unit #4 (in one of the darker buildings at the center of the image to the right of the three stacks for units #1 to #3) of 569 MW, and Unit #5 (on the top right in the other, darker building) of 860 MW. Both units #4 and #5 operate in a combined cycle.

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

Figure 2.13 Palo Seco Power Station, which uses Residual #6 (fuel oil) that is stored in the tanks shown in the upper portion of this image. This power station is equipped with two 85 MW and two 216 MW turbines for a total capacity of 602 MW.

the importance of economic factors in power grid operation. As discussed later in this book, economic conditions have an important impact on resilience because unfavorable economic conditions may limit investments, which, eventually, could otherwise have contributed to improve resilience. Such a relationship between economic conditions and resilience can be exemplified with the Puerto Rico Electric Power Authority's (PREPA) Palo Seco Power Station in Fig. 2.13. High costs of fuel for this and other, similar power plants was a contributing factor to PREPA's bankruptcy filing and associated reduced preparation for a disruptive event before Hurricane Maria affected the island in 2017. One alternative for thermal power plants producing steam to drive turbines is to burn coal, which is typically more economical than using fuel oil. However, even when storing coal at the power station is simpler than fuel oil or natural gas, a typical coal-fired power plant needs in many cases frequent and even daily coal deliveries for its operation, as exemplified in Fig. 2.14. Hence, the resilience issues associated to fuel delivery dependencies are still present in this type of power station. Also, ashes resulting from burning coal need to be properly disposed. Although fuel oil and coal-fired power stations, such as the one in Fig. 2.15, are being built, many of this type of power plant are being decommissioned due to high operation costs and negative environmental impact. Some of these decommissioned power stations were found near city centers, such as the iconic Battersea Power Station near the center of London and shown in Fig. 2.16, which had been replaced by power stations or renewable energy farms located away from the cities and thus the load centers.

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

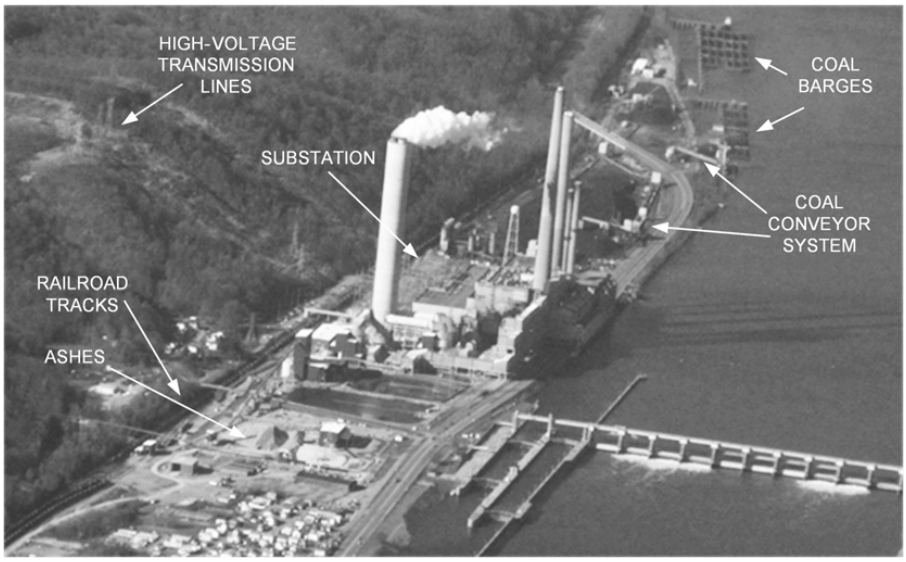

Figure 2.14 The W. H. Sammis Power Plant, which is equipped with coal-fired power generation units of the following capacities: four 170 MW units, one 275 MW unit, one 625 MW unit, and one 650 MW unit.

Figure 2.15 The Linkou Power Plant in Taiwan operating with two 800 MW coal-fired units while a third 800 MW unit was being constructed.

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

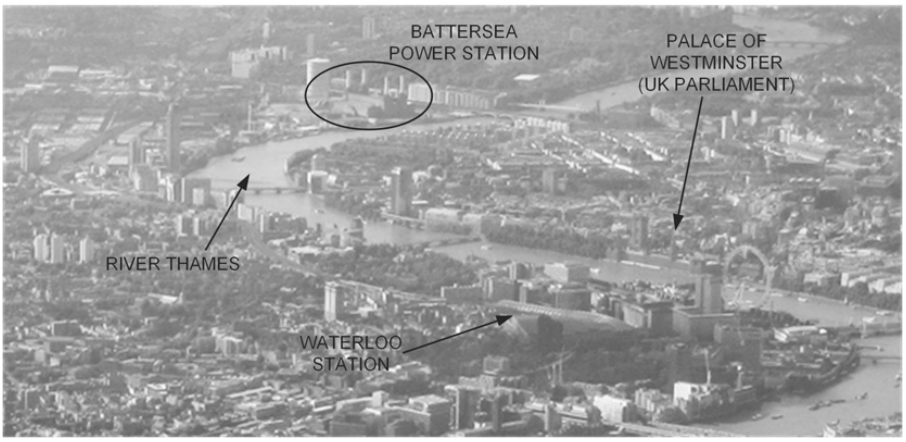

Figure 2.16 The decommissioned coal-fired Battersea Power Station located near the center of the city of London. Its maximum nameplate capacity was 500 MW.

In addition to depending on fuel delivery services, steam power plants depend on receiving water for cooling and to produce steam. Natural gas-fired power plants also depend on receiving water but only for cooling purposes. These water subsystems and other processes require provision of some relatively low electric power for pumps and other components. Hence power stations also have some dependence on electric power provision service. These dependences on both water and electric power provision services are a particularly critical need in nuclear power plants as exemplified due to the fact that both the nuclear accident in Chernobyl and the event in the Fukushima #1 power station were related to these dependencies. Additionally, nuclear power plants have a conditional dependence on diesel fuel delivery for their backup power generators as demonstrated by concerns for keeping cooling subsystems in Ukraine's nuclear power plants operating during the recent invasion by Russia. Because of their dependence on water supply, nuclear power stations are usually built next to a body of water or rivers, as exemplified by Fig. 2.17. However, such proximity to a body of water or rivers has associated risks, such as effects of floods or, as happened with the nuclear power station in Fig. 2.18, ice obstructing the water intakes from the River Loire. Both these images show the natural draft cooling towers, which are often identified with nuclear power plants. However, it is important to clarify for identification purposes within the context of Chapter 5 that nuclear power plants, particularly those in non-Western countries, may have different cooling towers with different shapes. Moreover, even in Western countries nuclear power plants may have cooling towers in different designs, as those in Fig. 2.19, or not even have cooling towers for those power plants, such as that in Fig. 2.20, using a water reservoir for cooling. Even more, natural draft cooling towers similar to those found in nuclear power plants are also found in other types of steam power plants. One difference between nuclear power plants and other thermal power plants is that although dependence on electric power for in-house equipment and on water supply is more critical in nuclear power plants because of the consequences if these needs are not satisfied, dependence on nuclear

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

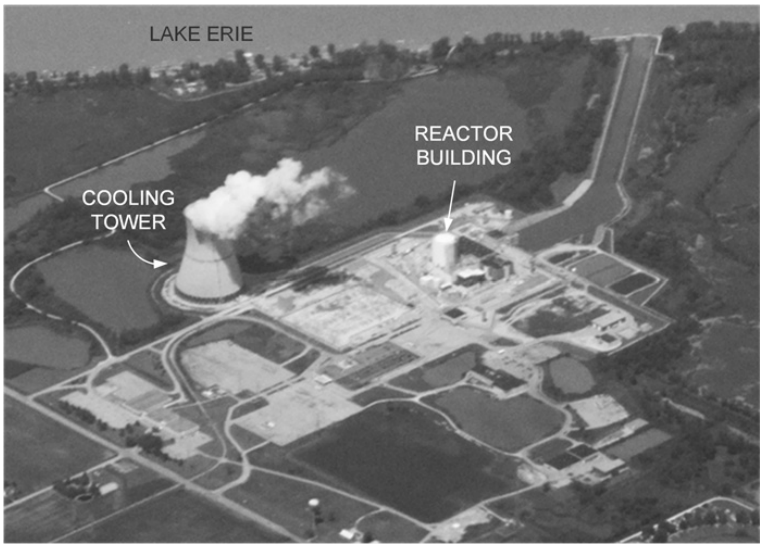

Figure 2.17 The Besse Nuclear Power Station, which has a capacity of 894 MW from its single reactor.

Figure 2.18 The Saint Laurent Nuclear Power Plant. The buildings for its two reactors, each producing 956 MW of electric power, are seen to the left of the image center.

fuel deliveries is less critical because nuclear fuel deliveries for a third of the reactor core typically occur every 12 to 24 months.

It is relevant to point out that there are other technologies that have been proposed for electric power generation, such as biomass (exemplified in Fig. 2.21), geothermal,

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

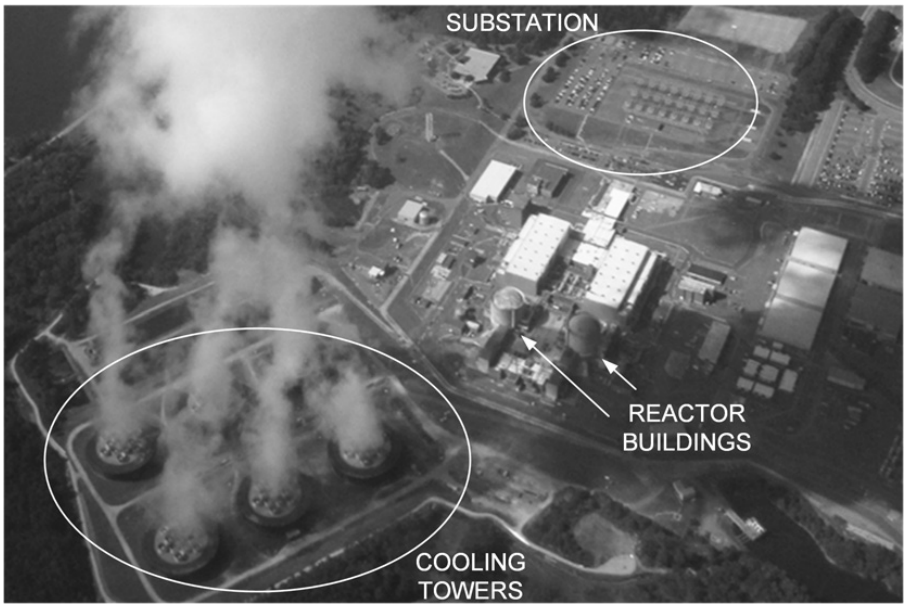

Figure 2.19 The Catawba Nuclear Station with its six mechanical draft cooling towers. Each of the two reactors of this nuclear power plant has a nameplate capacity of 1,155 MW.

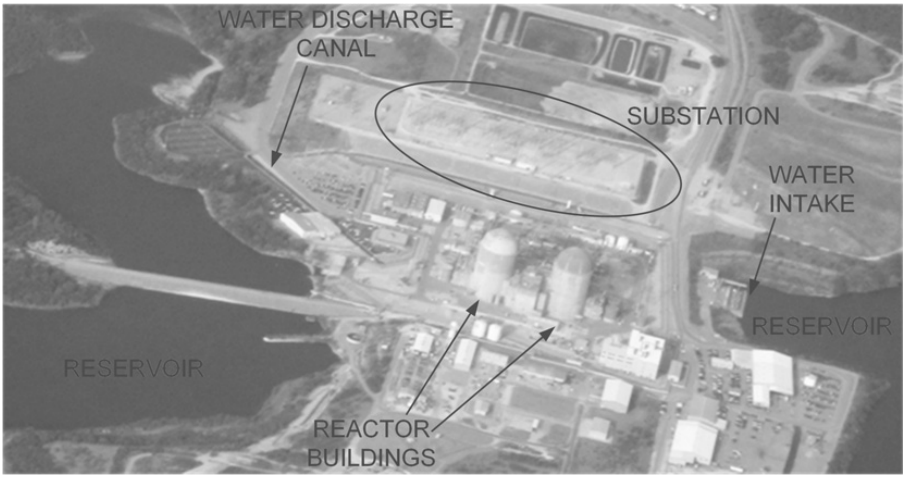

Figure 2.20 The Comanche Peak Nuclear Power Plant. It operates two reactors, each of them with a nameplate capacity of about 1,200 MW.

or ocean energy systems. However, these power plant technologies are not further discussed here because they have a small contribution to the total power generated in the power grids they connect to, with the exception of geothermal energy in Iceland, which represents about 27 percent of the power generated in the island. Other power generation technologies, such as fuel cells, are suitable for distributed generation applications, which are discussed in Chapter 6.

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

Figure 2.21 In the foreground: the HHKW Aubrugg AG power station, which uses biomass in the form of wood chips from the forests for fuel.

At an aggregate level, power grid loads typically follow a cycle with the lowest power consumption – namely, the base power consumption – at night, and with a higher power consumption – namely, the peak power consumption – observed during the day or evening, depending on the type of the dominant loads – namely, industrial, residential, or business – weather conditions, or other factors. Power stations need to be operated so that their power output equals the power being consumed by loads plus the power being dissipated in losses along the transmission and distribution circuits of the power grid. If power being generated exceeds power being consumed in loads and losses due to, for example, a sudden loss of loads, as happened during the February 2011 earthquake in Christchurch, New Zealand [1], the electrical frequency of the electric grid will increase. Hence, either power generators need to be commanded to reduce their power output to match the new power demand or, if the loss in load and losses is sufficiently large, then power generation units need to be taken offline. If, on the contrary, power generation falls below the power demanded by the loads and losses, either because of a drop in power generation or an increase in power consumption, then the electrical frequency of the power system will drop. Hence to prevent the frequency decreasing below the acceptable range, either power generators need to increase their power output or loads need to be intentionally shed. Although

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

changes in loads may happen relatively quickly, the rate at which frequency changes is usually less rapid because mechanical inertia present in the rotors of the generators powering the grid acts as an arrester for rapid frequency changes. Dynamic phenomena associated to frequency changes in conventional power grids have time constants typically ranging from a tenth of a second up to tens or a few hundred seconds, whereas time constants associated to boiler and other long-term dynamic phenomena are usually within a range of several seconds to a few minutes. Dynamic response can also serve as another way of classifying power stations. Power stations with a relatively slow dynamic response, such as coal-fired or nuclear generating stations, are classified as base-load power plants and they are operated at an approximately constant power output. Power stations classified as peaking power plants, such as those using natural gas turbines, have a faster dynamic response than base power plants, so their output is used to serve the peak load, which is observed for an hour or two during the day. Load-following power plants are used to power variations in the load and are more economical to operate than peaking power plants. The discussion in this paragraph shows that conventional grids have a top-to-bottom structure with a predominantly centralized architecture in which focus on planning and operation is first placed on a relatively few high-power stations – as compared to the many more much lower power loads.

Another inherent weakness to disruptive events found in conventional power grids is that although building a large interconnected system creates a system with more stable operating points, as more large generators add more "stiffness," such a large system also leads to long power transmission and distribution paths, which increase the chances of having disruptions along the path. Moreover, large systems are also more complex to operate, which also increases the chances of service disruptions caused either by external actions or simply by human error. An example of such complexity is found when calculating optimal power flows, which aim at identifying operational conditions by solving equations involving thousands of nodes and parameters with both physics constraints (e.g., circuit equations) and engineering constraints (e.g., power capacity limits of generators and transmission lines ampacity). Another example of the effects of such complexity arises when planning how to address contingencies in which multiple alternatives not only each require determining the new power flow conditions, but also may cause cascading failures that each become new contingencies.

In order to have high-power levels with reduced losses, high-voltage transmission lines are used to connect power stations to the power distribution infrastructure located close to the users. Although most high-voltage transmission lines are built overhead, high-voltage lines are also buried underground, commonly in urban areas to avoid safety, cost, and aesthetic issues of overhead high-voltage transmission in the dense urban environment, as demonstrated in Fig. 2.22. In the case of underground high-voltage lines, the dominant current technology is to use solid dielectric extruded cables with ethylene propylene rubber (EPR) or cross-linked polyethylene (XLPE) polymeric insulators (the latter used for higher voltages than the former). However, it is still possible to find oil-filled cables even when they are considered

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

Figure 2.22 Overhead high-voltage transmission lines in a dense urban area.

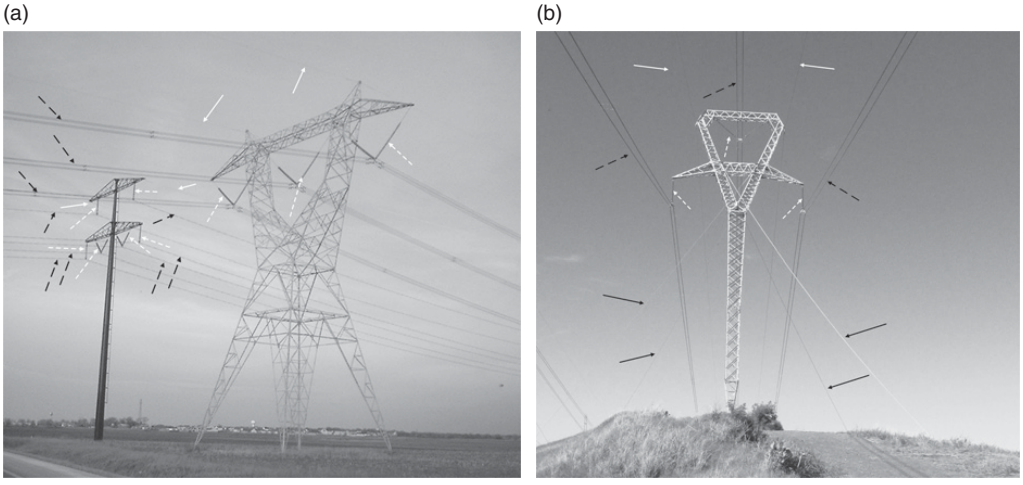

Figure 2.23 Examples of overhead high-voltage transmission towers, showing their main components: power conductors with dashed black arrows, shield wires with solid white arrows, insulators with dashed white arrows, and guy wires with solid black arrows. (a) Two self-supported towers, a metallic pole-type on the left and a lattice-type on the right. (b) A guyed tower.

obsolete. High-voltage lines are also sometimes laid down underwater, typically for connecting islands.

The most common approach to transmit electric power is with three-phase alternating current circuits operating at 50 Hz or 60 Hz, depending on the country. For this reason, high-voltage transmission circuits are usually identified by three conductors, shown in Fig. 2.23. Transmission voltage levels range from a few tens of kV, such as 39 kV, up to 765 kV, although lines with voltages below 100 kV are also considered to be subtransmission circuits. Depending on the country, common voltage levels are 66 kV, 69 kV, 132 kV, 138 kV, 220 kV, 330 kV, 345 kV, and 500 kV. In the case of overhead lines, these conductors are separated from the tower using insulators (also depicted in Fig. 2.23) made from porcelain, glass, or composite polymer materials,

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

Figure 2.24 Overhead transmission line using both wooden (darker colored) and metallic (lighter colored) poles. This line was damaged by Hurricane Ike.

which nowadays tend to be preferred over glass or porcelain because of their dielectric characteristics, lower cost, and for being less subject to being shot at in vandalism acts. It is also possible to find high-voltage direct current (HVDC) transmission lines, which are identified by their having two conductors per circuit. High-voltage dc transmission lines can provide a more flexible power flow control and have a lower installation cost than equivalent ac lines for distances sufficiently long in which the lower cost per kilometer of HVDC lines offsets the cost of ac–dc conversion stations at both ends of the HVDC line. Overhead high-voltage transmission lines also include one or two additional conductors placed on the top of and electrically connected to the tower or pole and then to ground. These conductors, called shield, earth, or ground wires, provide protection against lightning strikes, and they are also shown in Fig. 2.23. One technology for building high-voltage transmission line towers is using steel braces in a lattice structure, such as the one depicted in Fig. 2.23. Lattice towers are a common technology used for overhead high-voltage transmission lines because of their mechanical resistance, relative low cost, and installation ease. This type of tower can have a self-supporting structure, as shown in Fig. 2.23 (a), or they could be supported with the assistance of guy wires, as depicted in Fig. 2.23 (b). Concrete, wooden, or metallic poles, such as the ones in Figs. 2.23 (a) and 2.24, are another technology used for overhead high-voltage transmission line towers. Overhead high-voltage transmission towers can also be designed with a mechanical strength to support the weight of suspended conductors, like in the towers in Fig. 2.23, which are called suspension-type towers, or the towers can be designed so they can withstand the tension stress caused by the conductors. This latter type of tower, seen in the three rightmost towers in Fig. 2.22 and the towers in Fig. 2.25, is called dead-end or anchor towers, and they are

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

Figure 2.25 Anchor lattice towers in the port of Sendai, Japan. The tanks in the foreground were damaged during the tsunami of March 2011.

at the end of the transmission lines in substations or are used at regular intervals to interrupt cascading failures from falling towers, as observed in Fig. 2.24.

As Fig. 2.26 exemplifies, high-voltage transmission lines end on substations at both ends. Transformers at these substations step up voltages from generation power stations at the power transmission sending end, whereas transformers at substations at the receiving end (load-side) step down voltages to levels suitable for distribution circuits. Large transformers typically have power ratings above 100 MVA and are constructed so that each phase has a separate transformer, as exemplified in Fig. 2.27 (a). This figure also shows the firewalls separating each single-phase transformer to prevent fire propagating to other transformers in case one of them is ignited. Lower-power transformers are constructed with all three windings wound on the same magnetic core, as exemplified in Fig. 2.27 (b). This figure also depicts some of the main components of transformers. In particular, the bushings used to insulate the connection between the end of the connected transmission lines and the transformer windings are clearly seen in this figure. Bushings are usually a point of concern because of the mechanical stress they are subject to, especially during earthquakes if the connection of the bushings to the transmission line ends is not provided with sufficient slack to absorb changing spacing due to shaking. Another point of concern in transformers is all of the components related to oil handling, such as the oil expansion tank and the Bucholz relay, as shaking may make this relay open (oil is used in transformers for cooling, insulating, and preventing corona discharges and arcing). Oil leaks may also create hazardous conditions, and because oil is flammable, fire mitigation measures, such as the use of firewalls, are implemented when possible.

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

Figure 2.26 A transmission line end at a substation is seen at the bottom of this image.

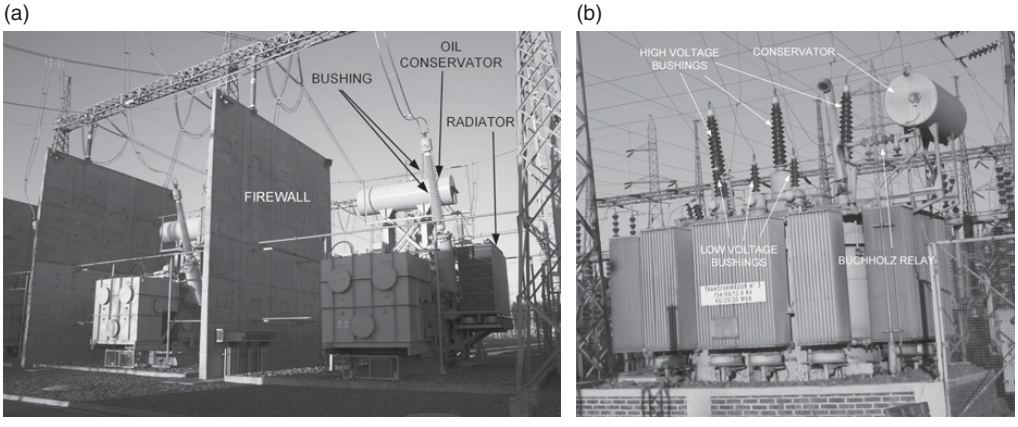

Figure 2.27 (a) Two single-phase 500 kV/220 kV 250 MVA transformers. (b) Two three-phase transformers seen in an aerial view.

Another important function of substations in addition to transforming voltages is to provide circuit protection and disconnect mechanisms with circuit breakers and other switchgear equipment. Figures 2.28 and 2.29 show an example of some of the typical circuit breakers found in high-voltage substations. The main function of these circuit breakers is to interrupt short-circuit currents when a fault occurs. Additionally, substations are equipped with disconnect switches, such as those exemplified in Figs. 2.29

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

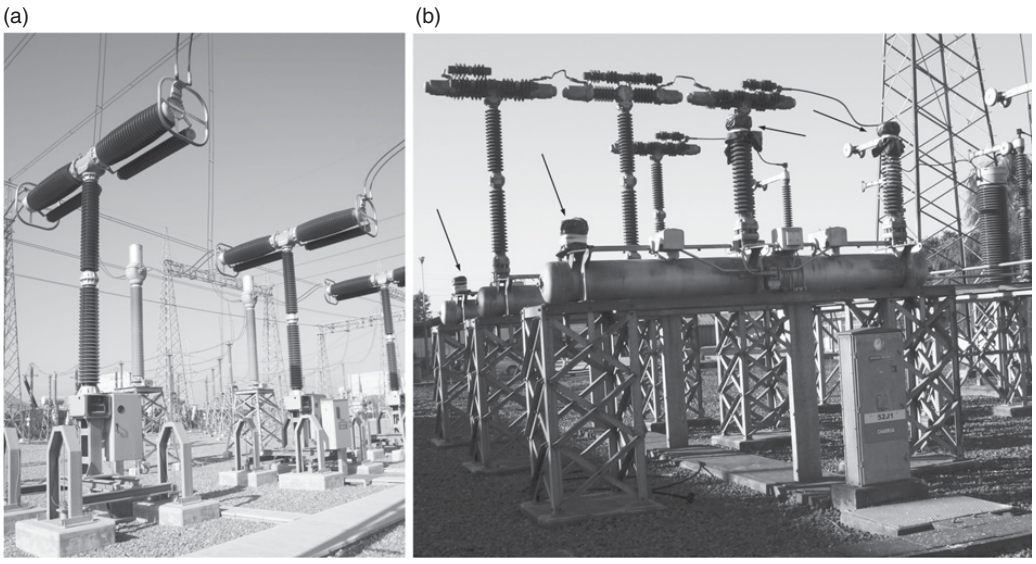

Figure 2.28 (a) Suspended dead-tank circuit breaker. (b) Air-blast circuit breaker with two of the three phases damaged during an earthquake (indicated by the arrows).

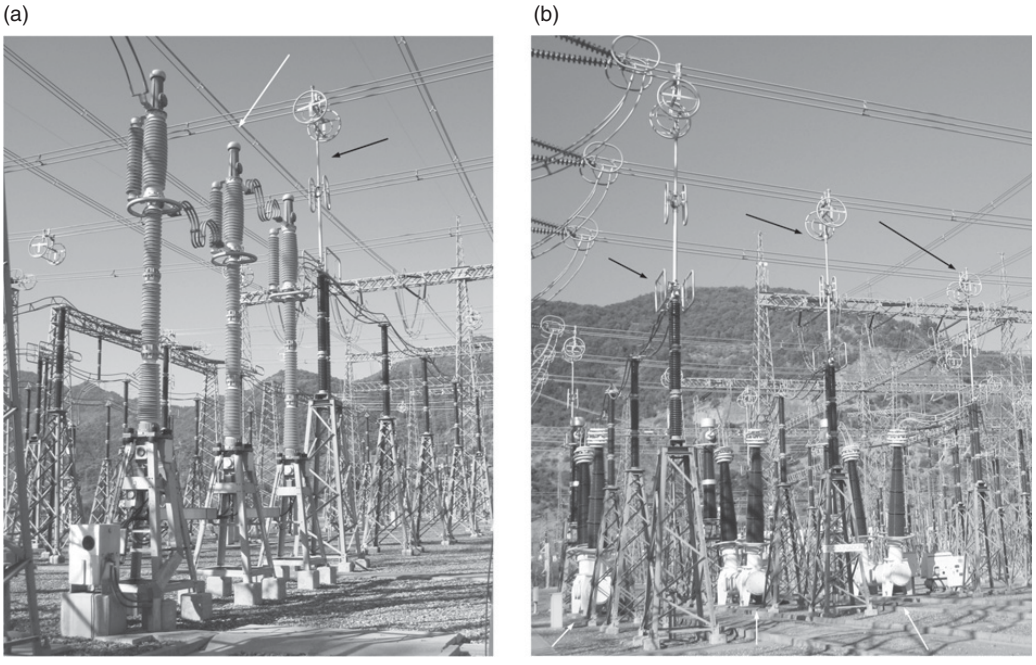

Figure 2.29 Example of circuit breakers indicated with white arrows and semi-pantograph disconnect switches marked with black arrows. (a) 500 kV candlestick live-tank circuit breaker. (b) Dead-tank sulfur hexafluoride (SF_6) circuit breaker.

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

Figure 2.30 Horizontal semi-pantograph (on the left) connected to current transformers (on the right).

and 2.30. These disconnect switches cannot handle high currents, so their main purpose is to provide a physical disconnection for electrical circuits.

In addition to circuit protection and disconnection and voltage transformation, substations serve other functions. Voltage is also regulated in substations by using capacitor banks (e.g., see Fig. 2.31), reactors, and other equipment for reactive power control. Voltage regulation also serves to control power flow along transmission lines. Sensing and monitoring is also performed in substations. Figure 2.32 (a) shows examples of potential transformers (PTs) and current transformers (CTs) used to measure voltages and currents, respectively. Recently, electric utilities have been deploying synchrono-phasors or phasor-measurement units (PMUs), which provide more advanced sensing capabilities than the PTs and CTs. Sensing and control signals from and to substations are usually communicated using dedicated networks that are part of the power grid and that are physically separated from publicly used wireline and wireless communication networks. An example of such power grid communication equipment is shown in Figs. 2.32 (b) and 2.33. Transmission substations also have power backup equipment, such as batteries and gensets (see Fig. 2.34), so the sensing, monitoring, and control subsystems can remain operating during power outages. Substations associated to HVDC systems also include power electronic conversion equipment, especially rectifiers and inverters, which are used to convert ac to dc or vice versa at the end of HVDC lines or to interconnect two ac systems, as is done in Japan to connect the grid operating at 50 Hz in the east of the country to the grid operating at 60 Hz in the west of the country. A similar conversion need also exists in the United

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

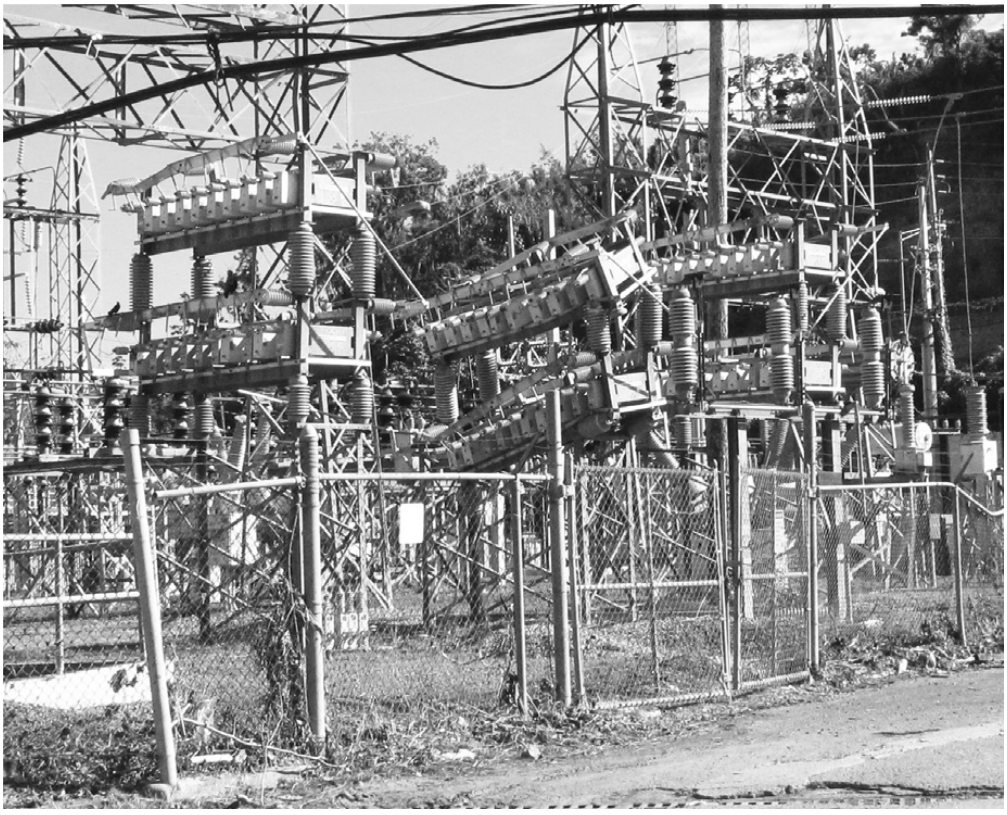

Figure 2.31 A damaged capacitor bank after Hurricane Maria.

States when interconnecting two of the three grids in the lower 48 states (the eastern interconnect, the western interconnect, and Texas' grid) because even when these three grids operate at a nominal frequency of 60 Hz, these frequencies may have minor deviations from their nominal value and the phase angle of the grids at the point of coupling could also be different, thus creating different instantaneous voltages on the interconnected systems.

Because of the critical nature of power generation units, transmission lines, and transformers, power grids are engineered with strategies to mitigate the already indicated weaknesses in terms of a top-down primarily centralized control, operations and planning approaches, and with very long electric power delivery paths. At the power generation levels, well-planned and operated power grids have reserve capacity. In substations, important high-power transformer banks have an $n + 1$ redundant configuration in which one extra transformer can quickly be connected to replace a failed transformer. In the case of single-phase transformers, the additional transformer serves as redundancy for a failed transformer serving a single phase. Where possible and economically practical, transmission line networks are built with a meshed architecture providing different paths for the power to flow. Additionally, some transmission lines may have 1 + 1 redundant configurations, or there could be

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

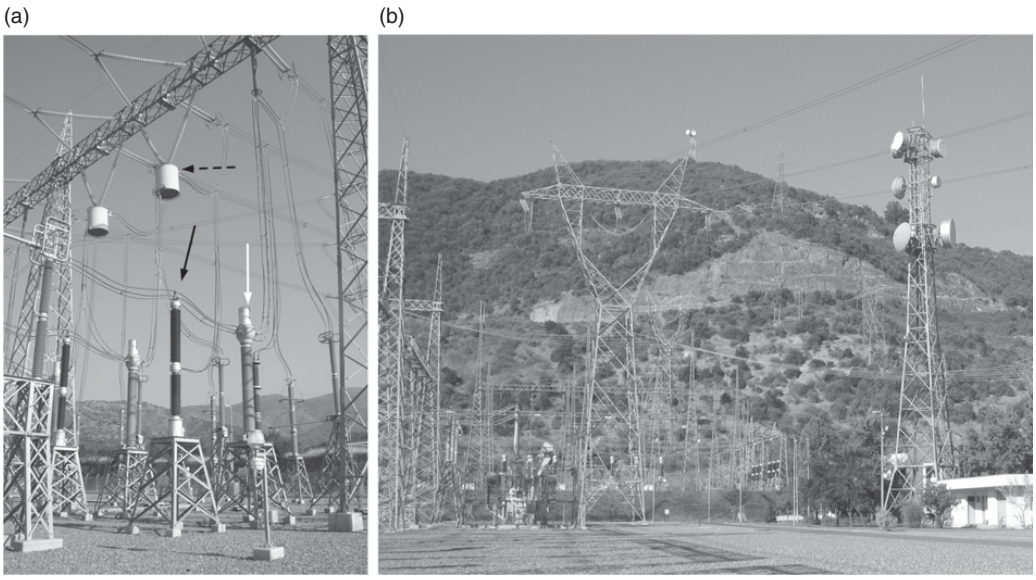

Figure 2.32 Sensing, monitoring, and control equipment in substations. (a) A wave-trap marked with a dashed black arrow used as part of a power line communication system, a CT marked with a white arrow and a PT marked with a black arrow. (b) Microwave antennae used for communicating to and from the substation in the image (notice the microwave antennae used as repeaters on top of the hill in the background).

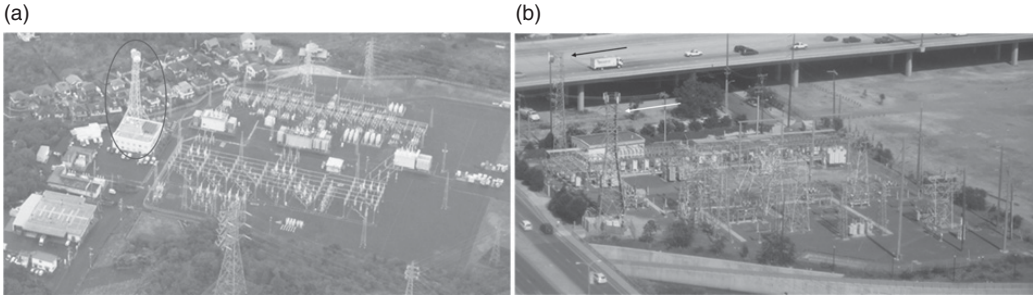

Figure 2.33 (a) Microwave communications antennae, which are part of the substation communications system. (b) A microwave antenna used as part of the substation dedicated communications network marked with a black arrow. Notice that the communications infrastructure for the substation is separate from the public wireless communications network, which uses a separate tower marked with a white arrow.

two parallel lines sharing their power transmission needs so in case one line fails, only half of the transmission capacity is lost. Moreover, in cases where it is practically feasible, transmission lines may be built with geographically diverse paths, so lines run at some distance from each other. Still, all these measures present a limited resilience improvement because power grids inherent already-indicated weaknesses persist. One such weakness is the practical near-impossibility of protecting the many kilometers of transmission lines and the many substations that are found in large interconnected

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

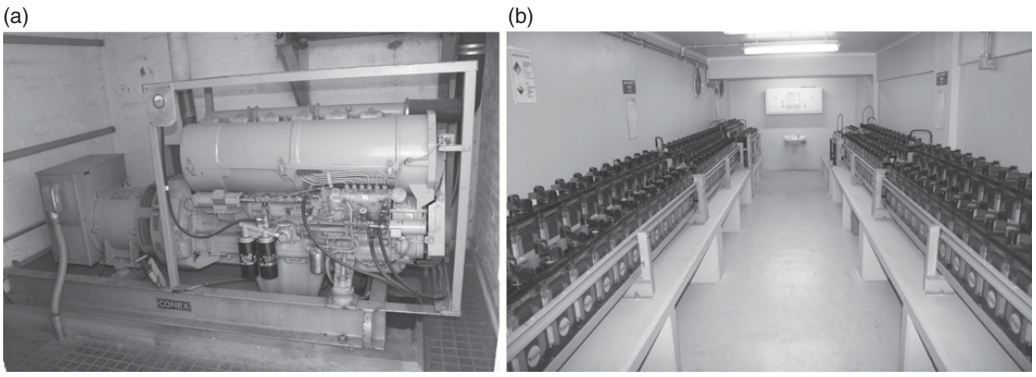

Figure 2.34 Power backup equipment in a substation. (a) Genset. (b) Batteries.

Figure 2.35 Aerial view of the Metcalf substation in California.

power grids, as exemplified by the attacks to transmission lines as a relatively common strategy used by insurgent forces during civil conflicts [2]–[3] or even in areas at peace as exemplified by the Pacific Gas and Electric Metcalf substation in Fig. 2.35, which was the subject of an attack in 2013, when gunmen damaged several transformers. Evidently, the buried transmission lines and underground substations often found in urban areas are more protected from both natural and man-made damaging actions, but their construction is more costly than that of overhead lines or aboveground substations. However, although locating substations aboveground in urban areas seems to be less costly than underground facilities, space constraints lead to compact substation designs, such as the one in Fig. 2.36, that may have a cost only slightly lower than that of an underground substation.

The power distribution portion is the part of an electric power grid closer to the consumers. As indicated, transmission lines deliver electric power to substations. At the substations, the voltage is stepped down with transformers, and the electrical energy is distributed to individual consumers located close to the substation, usually

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

Figure 2.36 Example of a compact aboveground substation in an urban area in Japan.

within a few tens of kilometers at most. Usually, electric power distribution circuits have a radial architecture in which there is a unique path from the substation to each of the consumers, as exemplified in Fig. 2.37. Still, in urban areas where different circuits could end close together, it is possible to find power distribution architectures that allow for laterals to be connected to alternative feeders, if necessary, thus creating a meshed or ringed power distribution architecture.

As Fig. 2.37 also shows, power distribution circuits start in distribution substations, such as those illustrated in Figs. 2.38 and 2.39, where there is at least one voltage transformation observed at distribution substations because the voltage output from a substation is usually 7.2 kV to 14.4 kV (medium-voltage), while the voltage at the ac mains drop of typical residential loads or other low-power loads – for example, small communication facilities, such as a cell site – is between 208 V and 480 V (low-voltage). The medium-voltage conductors at the output of a distribution substation are called feeders or primary feeders. Other medium-voltage conductors called laterals are then branched off from the feeder.

The final voltage step-down stage happens in small transformers placed along the lateral. Although there are different construction practices around the world, these smaller transformers are usually mounted on poles (see Fig. 2.40), concrete pads (see Fig. 2.41), or overhead platforms (see Fig. 2.42) and have power levels between 5 kVA and 200 kVA. Each of these types of transformers have advantages and disadvantages

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

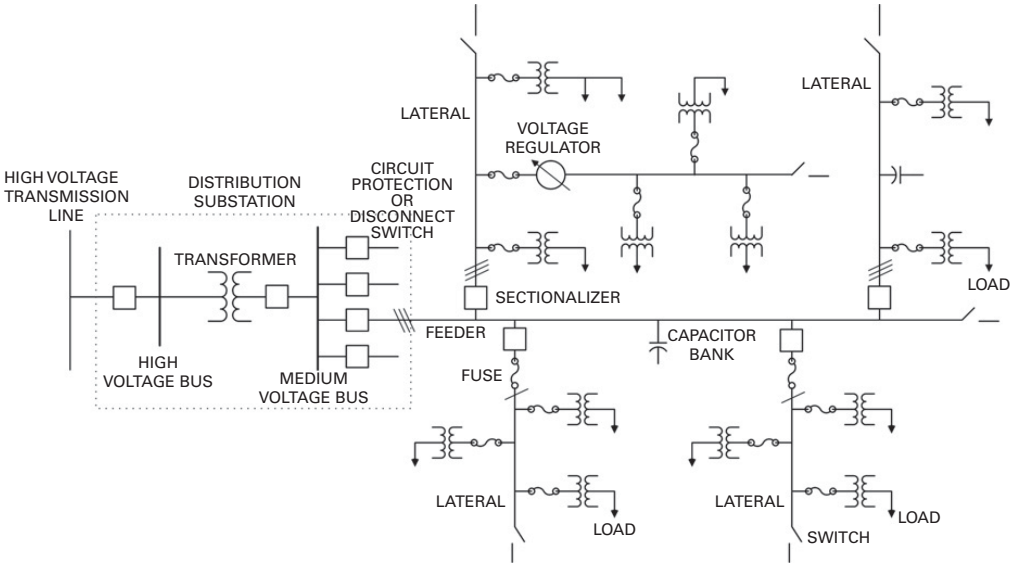

Figure 2.37 Example of a radial power distribution architecture.

Figure 2.38 A typical distribution substation in the United States.

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

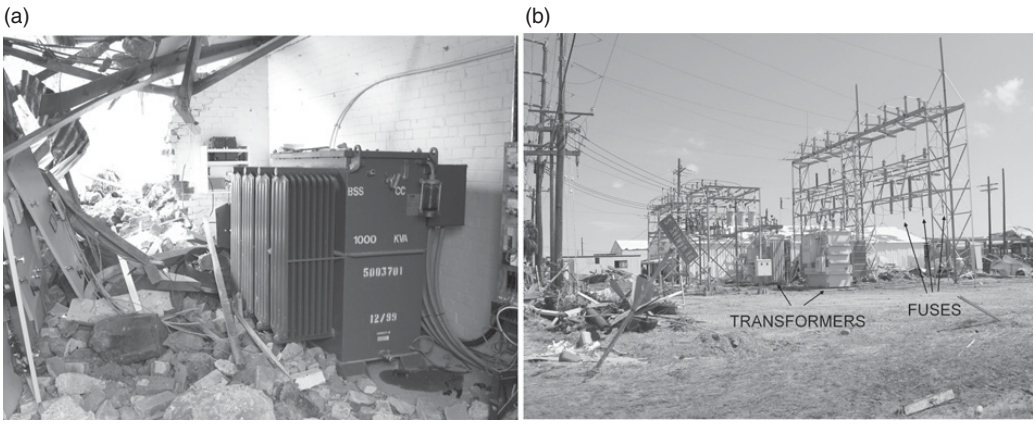

Figure 2.39 Damaged power distribution substations. (a) Substation located at the bottom of a cliff and damaged from rocks fallen during the February 2011 earthquake in Christchurch, New Zealand. (b) Substation damaged by Hurricane Ike.

Figure 2.40 Examples of pole-mounted transformers.

in terms of resilience. As Fig. 2.43 (a) illustrates, pad-mounted transformers tend to withstand storms but not floods, and have mixed resilience performance during earthquakes, as depicted in Fig. 2.43 (b). Overhead mounted transformers either directly on poles or on platforms tend to withstand earthquakes (see Fig. 2.44) and even tsunamis, although there are cases in which the shaking of the earth made the transformer fall to ground. Although overhead transformers tend to experience less damage with floods, their withstanding performance against storm surge is mixed, as shown in Figs. 2.45 and 2.46. Similarly, overhead transformers have a mixed damage performance during tsunamis, as exemplified in Fig. 2.47. Intense winds tend to cause considerable damage to overhead distribution systems, as illustrated in Fig. 2.48. In all these cases, damage to overhead transformers is mostly dependent on how well their

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

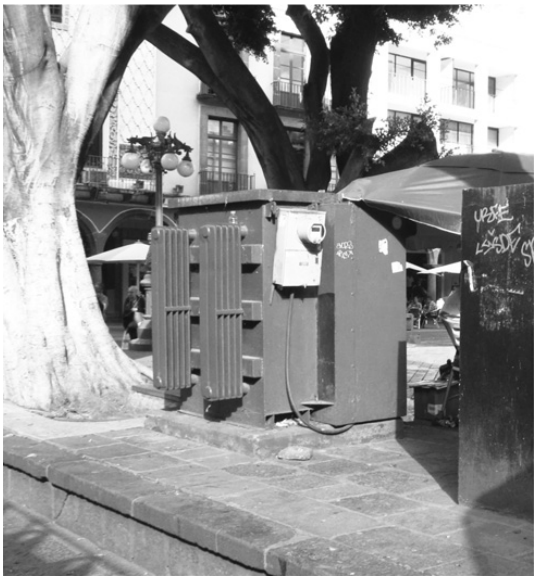

Figure 2.41 A pad-mounted transformer.

Figure 2.42 Overhead transformers mounted on platforms.

supporting poles withstand the damaging actions. The low-voltage output of these transformers is then connected to the loads either directly with a drop wire or through a short low-voltage secondary cable that connects the transformer to the drop from another pole or from a pad-mounted connection box located a few meters away from the transformer. In the case of higher power loads, such as medium-size communication central office or data centers with power consumption ranging from a few hundred kW, the conductor connecting to the customer premises may carry voltages in the

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

Figure 2.43 Pad-mounted transformers after natural disasters. (a) After an earthquake. (b) After a tornado.

Figure 2.44 Overhead transformers on platforms after earthquakes and tsunamis.

range of 1.2 kV to 4.2 kV, and thus requires additional transformation stages inside the consumer facility. Larger loads, on the order of a few MW, have electric power connections at medium voltage levels.

In addition to transformers, power distribution circuits have other components. Besides fuses and disconnect switches, these circuits include voltage regulators, shown in Fig. 2.49 (a), which are inductors that can be adjusted depending on the voltage levels along the line. Similarly, power distribution circuits include capacitor banks, such as that in Fig. 2.49 (b), which also affect voltage profile by allowing one to adjust their corresponding circuit power factor.

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

Figure 2.45 Pole-mounted transformers after suffering a storm surge. (a) Surviving transformers. (b) A damaged transformer but still mounted on the pole.

Figure 2.46 Damaged transformers after a hurricane. The one shown in (a) was damaged by pole failure.

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

Figure 2.47 Pole-mounted transformers after an earthquake and tsunami. (a) Transformer damaged due to a fallen pole. (b) Undamaged transformers.

Figure 2.48 Transformers damaged due to strong winds.

Arguably, one of the most important technology changes observed since the first electric grids were developed more than a century ago is in their loads. Initially, the main loads were lights, induction motors for industries, and to a lesser degree traction motors for transportation applications. At this time, the end of the nineteenth century, there was a need for electrification solutions able to power induction motors. This is one of the reasons why three-phase ac power grids prevailed over dc systems and thus is a contributing factor for the design we observe nowadays in large, interconnected power grids. Another important contributing factor for such a design is the need to provide electricity to the largest number of consumers at the lowest possible cost and with sufficient reliability. Yet notice that resilient operation was not at the time a design goal for electric power grids as the main technological solution for electrifying societies. However, modern-day economies experiencing an electronic

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

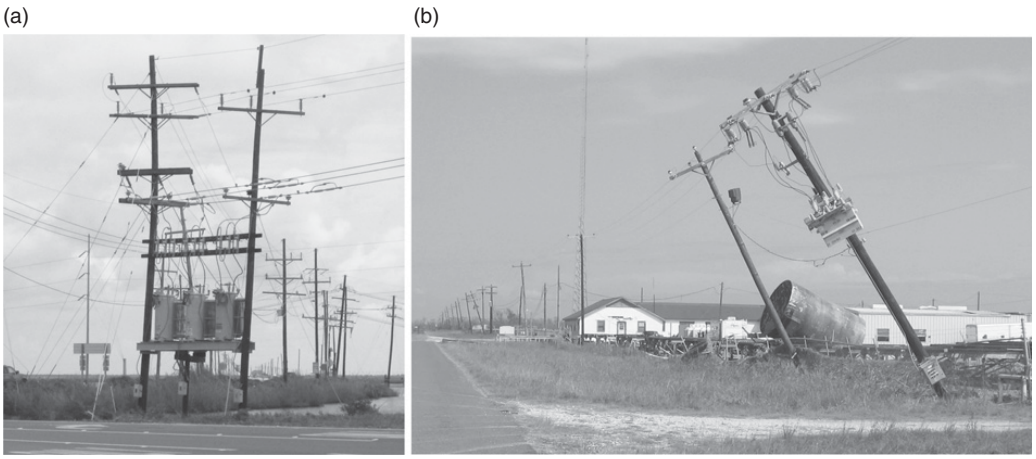

Figure 2.49 (a) Voltage regulators. (b) A pole-mounted capacitor bank (the pole is tilted due to high winds and storm surge from a hurricane).

and information revolution are moving toward being less industry based and more service oriented. Thus, nowadays electronic loads, such as computers, become a main driver for planning, designing, and operating electric power systems. Moreover, as business operations, banking, and finance are already significantly based on electronic transactions and less on paper cash, economies are becoming increasingly integrated to information and communication technologies (ICT) systems. Thus, ICN facilities are an extremely critical load, which are added to already existing critical loads, such as hospital and safety or security enforcement facilities. These loads nowadays have resilience and reliability design requirements for electrification that traditional power grids are not able to provide. One of the issues that ICT loads present to power grids in order to achieve a resilient power supply is that these loads are widely distributed and most of them – namely cell sites or broadband outside plant systems – have relatively low power consumption, similar to homes. Thus, contrary to critical loads in buildings, such as hospitals, these loads are connected to regular circuits to which are connected noncritical loads, such as residences, businesses, or small- to medium-size industries. Hence, in these conditions it is not possible to provide to these loads a resilience level different from that of the noncritical loads connected to the same distribution circuit. Moreover, since ICT service users interact with ICNs through devices, such as smart phones or computers, that require electric power for their operation (or at least, in the case of smart phones, tablet computers, or laptops, need their batteries to be recharged somewhat often, such as once a day), the need for resilient power extends to residences and businesses that are not considered critical loads. Similar concerns will also become more prevalent as the use of electric vehicles becomes more extensive, because without the means to recharge their batteries, people will see their mobility significantly reduced in the aftermath of a disruptive event, which is an important need under such conditions.

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

Figure 2.50 A fallen pole with transformers causing loss of power to consumers connected to this transformer. Notice that electric power service is on across the street.

When compared to failures at the generation or power transmission levels, a failure in a distribution circuit almost always affects fewer consumers, and in cases such as that exemplified in Fig. 2.50, neighboring loads having an outage due to nearby equipment damage may not experience loss of service. Thus, restoration activities for power distribution circuits may have a lower priority than for generation and transmission components. Furthermore, failures at the distribution level are typically more numerous and more geographically distributed than those at the transmission and generation levels. Moreover, the relatively common scenario of restoring service in power distribution circuits run with multiple overhead drops and conductors, as illustrated in Fig. 2.51, tends to be tedious work. All these reasons contribute to longer restoration times for power distribution circuits even when failures at generation and transmission levels are usually more complex to repair than at the distribution level.

During the past two decades, power grids have been adopting technologies, such as smart meters, PMUs, and demand-response energy management control systems, that are part of what are called "smart" grid technologies. These technologies increase communications, control, and sensing capabilities with respect to how conventional grids had been monitored and operated. Depending on how these technologies are used, they could have a beneficial effect not only on power grid resilience but also on ICN resilience. Thus these technologies are further described and discussed in Chapters 6 and 8 of this book, with a description about ICNs presented in the following section.

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

Figure 2.51 Example of a power distribution circuit with a dense run of conductors (some of them could also be telephony cables).

2.2 Information and Communication Networks Design and Operations Fundamentals

2.2.1 Concept of Communications Networks, Protocol Layers, and Packet Switching

In the context of information and communication technologies, a network can be broadly seen as a collection of electronic devices, called "nodes," that are interconnected to exchange data. The network nodes are classified as either end system nodes, which are the ones at the ends of the network, or switching nodes, which form the mesh-like structure of the network by interconnecting each other and end systems. End systems are very varied in nature and include, for example, computers, web servers, sensors, cellular phones (in a voice or data connection), gaming devices, or satellite radios. These examples of end systems illustrate how many of them provide the interface for end users to connect and interact with other end systems in the network. The connection between two nodes is called a "link."

Historically, modern communication networks evolved from the first communication system based on transmitting electrical signals: the telegraph network. The telephone network in its traditional form, what is called the "Plain Old Telephone Service" (POTS), developed from the desire to communicate human voice over the cabling of the telegraph network. Because of this, the POTS network inherited from the telegraph system a "circuit switched" form of connecting and exchanging information (human voice) between end systems (the telephones). In circuit switching, the communicating end systems undergo a procedure to establish a connection that entails the

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

assignment and reservation of the network resources necessary for the end systems to communicate. In the POTS case, this procedure took place during dialing the called telephone number at the calling telephone, and the resources being assigned were actual circuits connecting from the calling telephone, going through intermediate switches, and ending at the called telephone. The differentiating characteristic of circuit switching is that the network resources being assigned during the establishment of the call are reserved for the exclusive use of the call for as long as it lasts. This model of network operation based on circuit switching was the preeminent paradigm for roughly one hundred years, from the invention of the telephone to the late 1980s when, following the progress in computing, networks based on the different "packet switching" paradigm began to gain ground. At the time of gradual transition from circuit- to packet-switching technology, the motivation in this change was the inherent inefficiency associated with circuit switching's exclusive reservation of network resources for each call, instead of dynamically sharing them among multiple calls, as is the case with packet switching. Today, most communication networks are based on the packet switching paradigm because it has the highest efficiency and it fits much better with end systems that generate and process information in digital form (as is the case with most of today's end systems).

An overview of packet switching networks is presented in the few next paragraphs. But before doing so, it is worthwhile to consider another more abstract, but quite insightful, definition for an ICT network as an infrastructure that allows running distributed applications across computing systems. A good example of a distributed application is web navigation, where one part of the application is the client web browser software and the other part is a web server. In this example, when a user enters a web address in the browser and hits the Enter key, the client proceeds to send a request for data (a web page) to the web server at the entered address. When the server receives the request, it proceeds to retrieve the requested data from its memory and sends it back to the client computer running the client web browser, which proceeds to present the data in a format for the end user to read it. The application in this example is distributed because the client and server components are generally separated by large distances, yet the end user does not experience directly the distributed nature of the application thanks to the infrastructure of the network. (From the end user's perspective, the web browsing application could be running on its local computer and still accomplishing the same action of retrieving a specific piece of information.) Note that in this example, as is always the case, the network infrastructure is formed by both hardware and software components. In particular, the main physical (hardware) core components of the Internet as a data network are datacenters, which are the facilities where the web servers are located. Datacenters contain many servers, sometimes hundreds of them, to provide a variety of data services, such as web hosting. Because of the large number of servers, large datacenters, such as the one in Fig. 2.52, have a large power consumption, on the order of a few megawatts, 30 to 40 percent of which is consumed by the air conditioning system needed to dissipate the heat created by the servers. The Internet is thus formed by interconnected datacenters in a meshed network of the different Internet service providers (ISPs), which connect at exchange points, such as the one in Fig. 2.53, which

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

Figure 2.52 Google's data center in Henderson, Nevada. The main electric substation is seen on the top left and a secondary substation is observed between the two buildings. Cooling and power backup equipment is seen between the two buildings and on the right.

Figure 2.53 The building of the NAP (network access point) of the Americas.

interconnects data traffic among 150 countries around the world, primarily in the American continent. These networks typically have a backbone used to connect the facilities with the largest data capacity in terms of data transmission and processing. Moving toward the edges, each ISP may connect to the web clients through regional ISPs. The Internet has also a hierarchical and decentralized software network architecture used to resolve the addresses; that is, to associate a web address initiated by a web client to the Internet protocol (IP) address of the server where the desired web page resides. Resolving the address is a function performed by so-called domain name servers (DNS), which are located in data centers that are critically important for the operation of the Internet because issues with DNS would prevent a connection being established between the web client and the specific intended web server.

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

In order to run distributed applications effectively, the enabling infrastructure (the network) introduced in the previous paragraph needs to implement many functions. These functions not only enable the exchange of information between two nodes through a link, but also realize the forwarding of a message across multiple nodes until reaching the intended destination. Since the early times of networking, it was realized that the best approach to organize the many functions was through a modular architecture that groups the functions based on the scope of what they achieve. In this way, a group includes all the functions that are needed to convert individual bits into electrical signals that are subsequently transmitted, and another group includes all the functions needed to route information by following a path formed by multiple links. Conceptually, these groups are organized into a stack, called the "network protocol stack," where each group forms one layer of the stack. By convention, the layers in the protocol are organized from the layer that implements the most fundamental functions necessary for communication at the bottom of the stack and proceeding toward the layer at the top of the stack, building up increasingly complex networking function. Each layer in the stack uses the services (the functions) implemented by lower layers. Data flow only between neighboring layers.

Figure 2.54 shows the five layers that form the Internet Protocol stack (also known as the TCP/IP protocol stack) and the function of each of the layers. From the bottom of the stack to the top, the five layers are as follows:

– *Physical Layer:* The first layer, at the base of the stack, called the "Physical Layer" (or often also simply called "Layer 1" or the short form of "PHY" layer), provides

Layer	Function
APPLICATION	Provides the interface for applications to send information over the network.
TRANSPORT	Provides for the functions to manage sessions between end systems and for the reliable end-to-end (whole path over the network) transfer of the sequence of packets from the session.
NETWORK	Provides for the functions to forward packets to the intended destination through a network path formed by multiple links.
DATA LINK	Provides for the reliable transfer over a link of information bits, organized in a frame, and for communication channel access arbitration.
PHYSICAL	Transmits unstructured bits over a communication channel.

Figure 2.54 Internet protocol stack.

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

the functions that convert bits into electrical signals or electromagnetic waves, in a process that is called "modulation." The definition of a Physical Layer involves not only the specification of how to convert bits to electrical signals, but it also includes other related specifications as, for example, the duration and timing of signals, power levels, and spectral characteristics that the signals are expected to follow. The layer above the Physical Layer gradually builds up functions to completely implement a network.

– *Data Link Layer:* The second layer, which sits directly above the Physical Layer, is called the "Data Link Layer." This layer deals with functions needed to directly connect two nodes, forming what is called a "link." As such, this layer is arguably the one that covers the broadest set of functions, including how to organize the bits to be transmitted into a structure called a "frame," how to deal with the inevitable bits that will be received in error after transmission (by implementing procedures to detect errors and correct them at the receiver when possible or otherwise managing the retransmission of the frame with errors), and how to arbitrate access to an often shared transmission medium between multiple transmitting nodes. A node that has the hardware and software to implement the functions of the Physical and of the Data Link Layer can connect to another node in a "Point-to-Point" communication. However, it lacks the capability to form part of network (in the sense of communicating over a path formed by connecting multiple links). Because wireless networks often follow a topology formed by point-to-point connections, it is very common to see that the standards that define any given wireless networking technology usually address only the functions of the Physical and the Data Link Layer. It is because of this also that the functions of the Physical and of the Data Link Layer are typically implemented in a single microchip that forms the central component of the network interface card (NIC) at a node.
– *Network Layer:* The third layer, called the "Network Layer," provides functions to establish a path between a source and a destination node that is formed by the concatenation of potentially multiple links, and to route information through this route. As such, this layer is the first (when going from the bottom to the top of the network protocol stack) that implements functions needed for a node to operate in a network. Typical functions that pertain to this layer are the definition of addresses for the nodes (how to uniquely identify each node in the network) and the procedures to announce and/or discover routes over the network.
– *Transport Layer:* The fourth layer, the "Transport Layer," deals with addressing reliability issues that result in errors or data loss when information travels through the network. An end node that implements the functions of the Network Layer becomes capable of sending data through a path comprising multiple intermediate nodes (and possibly multiple different intermediate networks, as is the case with the Internet). In these cases, the rate at which the source node is sending packets may prove to be too large for an intermediate node that may be experiencing a high volume of traffic, thus driving the intermediate node into a congested state. In addition, with communication over a network, a sequence of packets may arrive at the destination in an order different from the one they left the source node (because

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

each packet may potentially follow a different route). Therefore, the functions implemented by the Transport Layer are intended to deal with these issues.

– *Application Layer:* The layer at the top of the network protocol stack, the "Application Layer," provides the interface that connects applications to the network protocol stack. Because of its function, a node usually instantiates many protocols from the Application Layer, each being specific to a particular application. For example, a node may instantiate a Hyper-Text Transport Protocol (HTTP) that is used by web browsing client applications to form packets of information to query for the file associated with a webpage at some server on the Internet and, further, to interpret and process the data received in response to the query. In some cases, an application may instantiate multiple protocols from the Application Layer. For example, an email client application may create an instance of the Simple Mail Transfer Protocol (SMTP) to send email messages and also create instances of the Internet Message Access Protocol (IMAP) and of the Post Office Protocol (POP) to receive email messages.

Recalling that a network can be seen as the infrastructure to enable running distributed applications, it is interesting to note that, with the exception of the Physical Layer, the protocols at each layer of the stack operate as distributed applications themselves. To do this, at the data source node extra data needed by the protocols is appended to the data from an application as it is passed from one layer to the next. As Fig. 2.55 exemplifies, the extra data that is added at each layer is used by the portion of the protocol corresponding to the same layer that is running as a distributed application on the receiver side. For example, the protocol at the network layer would append as extra data the address of the destination node so that nodes that receive this extra data along the path to the destination will read and interpret it at their network layer protocol software and proceed to route the information accordingly. In another example, the transmitting node would add at the Data Link Layer extra data that will allow the

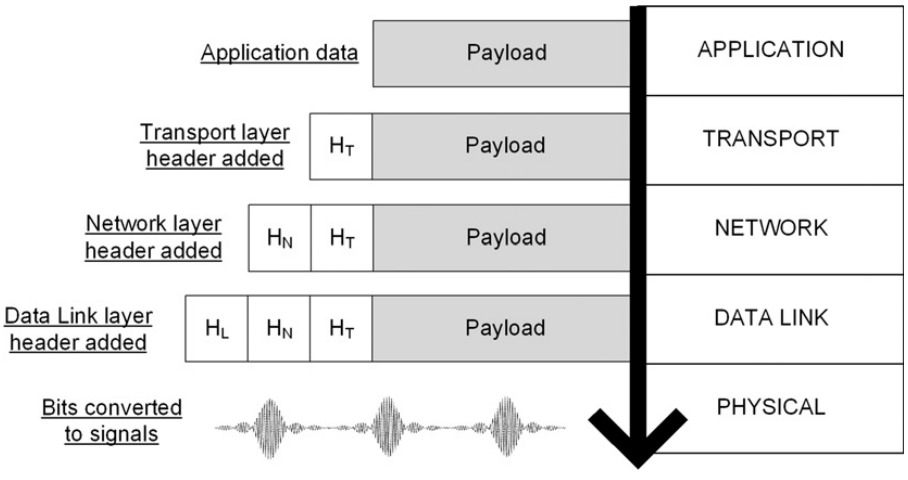

Figure 2.55 Packets over in the protocol stack.

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

receiving node to check if there were errors that had been introduced to the data bits during transmission over the link. Therefore, the process followed for the transmission of a block of data involves at each layer the addition of extra "data about the data" (what is literally called "metadata"), which is all transmitted as a block of bits generically called a "packet." As such, a data packet is a structure composed by a payload (the data from the application) and the metadata, which is placed in a substructure called the "header." The packet header itself can be divided into multiple headers, one for each of the protocol stack layers except the Physical Layer, each containing the data necessary for the protocol of the corresponding layer to perform its function.

The creation of the data structure called a "packet" not only was a necessity to implement protocols running as distributed applications, but it also enabled a new form of network called a "Store-and-Forward" or "Packet Switched" network. Because the packet carries the payload and all the data necessary for the payload to go through the network without the need of any other side information, the packet can be stored in any node along the network path until the node can process the information in the header and continue the forwarding of the packet along its path on the network.

2.2.2 Point-to-Point Communication

As explained earlier in this section, one of the most important elements in establishing a communication network is the forming of links between two nodes. The direct connection between a transmitter and a receiver through a link is called a point-to-point (P2P) communication. Figure 2.56 shows a simplified block diagram for the transmitter in a P2P connection and provides an overview of the main operations that are performed in order to transmit data. In the diagram it can be seen that the information source (for example, speech, video, or a file from a computer) is input into a "source encoder." The function of the source encoder is to convert the analog or digital information source into a sequence of binary digits. When the information source is analog, this process of representing the information from the source into a bit stream has as a first step an analog-to-digital conversion, after which the information is in digital form, as is already the case for a digital information case. From this point, the process followed is of quantization and, often, source compression. The source compression operation is intended to reduce the amount of bits to be transmitted and it can be of the type of lossless compression (when the original information source can be

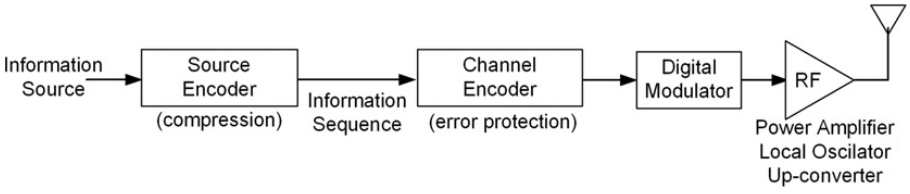

Figure 2.56 Simplified block diagram for a wireless transmitter.

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

recovered in its exact original form after performing a decompression operation) or lossy compression (when some information from the source is irremediably lost in the process of compression and decompression).

The source encoder output, called an information sequence, is passed through a channel encoder. The channel encoder introduces into the information sequence extra, redundant bits that are called *redundancy*. The redundancy is introduced in a controlled way so that the information and the redundant bits are interrelated through a deliberate structure. The purpose of introducing redundancy in such a controlled way is to enable at the receiver the recovery from the errors that have occurred during the communication process. The particular structure with which the information sequence has been modified with the introduction of redundancy allows for the detection and correction of the errors occurring during transmission. As a broad and general rule, the larger the proportion of redundant bits within the transmitted data, the larger the number of transmission errors that can be detected and corrected. However, introduction of more redundant bits comes at the cost of requiring a larger bandwidth or transmission capacity for communication of the bit stream at the output of the channel encoder.

Figure 2.56 shows that the output of the channel encoder is fed into the digital modulator. The purpose of the modulator is to convert the binary information sequence into an electrical signal, called the modulated signal. As such, this operation converts digital bits into an analog signal. The signal at the output of the modulator is oscillatory, with the information that is contained within the input bits mapped into one or a combination of the amplitude, frequency, or phase of the modulated signal. Following the digital modulator, the modulated signal's power is amplified and the main frequency of its oscillation is increased to the one needed for transmission. After this, the resulting signal is fed to the antennae elements or to a cable from where it propagates through the transmission medium until reaching the receiver.

A communication channel is the entity over which the transmitted signal propagates from the transmitter to the receiver. During propagation over the channel, the transmitted signal may be affected by different physical phenomena. Some of the most common phenomena are signal attenuation, addition of interference, filtering of some frequency components, and addition of noise. Noise is usually modeled as being added to the transmitted signal in the channel, although in reality it originates from thermal noise associated with the electrical and electronic components of the receiving circuit. These phenomena introduce impairments to the transmitted signal, which, after processing at the receiver, may result in the introduction of errors into the transmitted bit stream. The rate at which these errors occur is called the "bit error rate" (BER) and is, in effect, the empirical probability of occurrence for a bit error. The error rate can also be measured in other related quantities depending on the communication protocol layer that is considered. In this way, the error rate at the Data Link Layer is often called the "frame error rate" and at the Network Layer it is called the "packet error rate."

The physical characteristics of the channel determine the maximum rate of information that can be transmitted while remaining feasible to control the error rate. This maximum rate is called the *channel capacity* and is measured in units of bits

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

per second. The channel capacity as just defined is a theoretical quantity calculated under idealized settings. The actual rate of information that is transmitted over a link is called the *throughput* and is also measured in units of bits per second.

2.2.3 Wireline Telephone Networks including Those Used for Plain Old Telephone Services

Wireline telephone networks are the aforementioned traditional landline telecommunication system in which subscribers with fixed telephones are interconnected through cables by a commutating element called the switch. Other names given to this network are the public switch telephone network (PSTN), fixed-telephone network and, as mentioned, POTS network. In the United States, companies that provide POTS services are called a local exchange carrier (LEC).

A switch is the most important element of a PSTN. The function of a switch is to link two subscribers to the network by commutating calls. The building that houses the switch is called a central office (CO), which represents the core network facility. In some ways, a CO is analogous to a substation in an electric utility grid, serving as the primary connection point to consumers as well as an interconnection to bulk communication infrastructures. Each CO covers a portion of the LEC territory. This geographical region is called a CO area and is determined based on the number of potential subscribers (i.e., the users), geographic limitations, and demographic characteristics. Usually, a CO is located in the most populated zone within its area to minimize connection length to subscribers. At the same time, the location of the CO must be close to important routes to minimize linkage distance to other COs.

The CO also contains other communication equipment, including transmission systems necessary to connect the CO to other COs. Figure 2.57 shows the basic communication elements of a CO. Calls are electronically commutated in the switch matrix, which today is realized with computers[1]. Cables to the subscribers are terminated in vertical blocks in the main distribution frame (MDF). These cable terminations are connected with cross-connect jumpers to horizontal terminal blocks that are also located in the MDF. Several positions in the horizontal blocks of the MDF are then connected to switch line modules (LM), where the signals of some LMs are combined in multiplexing units (MUX) to produce a single signal in order to reduce the switch matrix complexity. The multiplexed signals are processed in interface modules (IM) that separate the MUX units and the switch matrix. The switch matrix is also connected through IM and de-MUX units to trunk modules, where high-capacity links are terminated in the switch. These trunks are then connected through the transmission distribution frame (TDF) to the transmission system, where most of the trunk signals are routed to other COs and communication centers. The entire system is controlled with process controllers and managed from administration terminals. In more modern

[1] In the first telephone networks dating about 100 years ago, human operators were in charge to connect the calls. Later, human operators were replaced by automated systems using electric relays. Transition to the modern computer-based digital switch systems started in the 1970s.

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

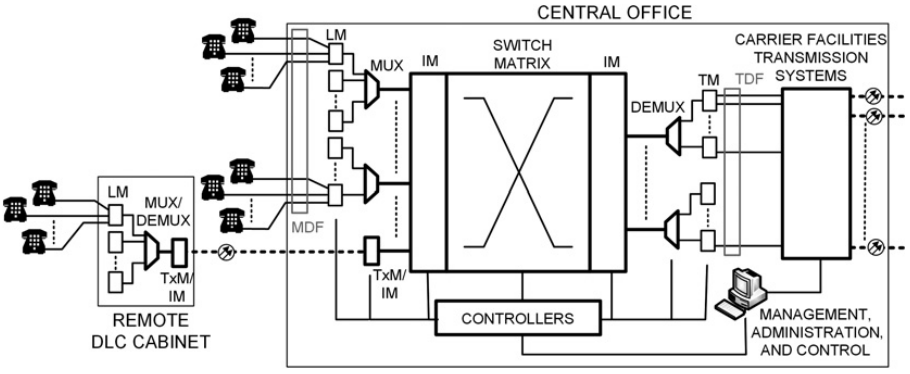

Figure 2.57 Central office main communication components with a remote digital loop carrier (DLC) system.

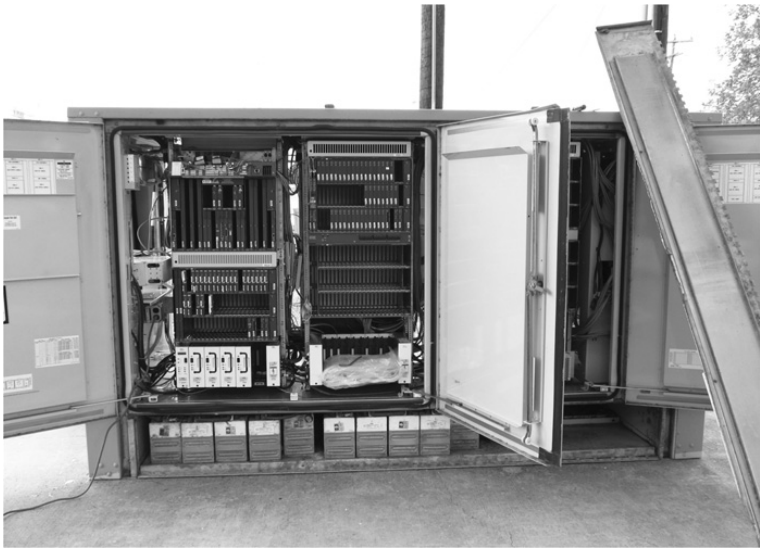

Figure 2.58 An open DLC cabinet. The batteries are placed at the bottom to prevent them from receiving heat from the operating equipment.

systems, the architecture of CO resembles more that of a data center in which connections among the servers are made with fiber optic cables and communications are established with packet switched protocols.

Figure 2.58 also shows an alternative to connect subscribers through a multiplexed remote terminal: a digital loop carrier (DLC) system. The DLC remote terminal is placed away from the CO in metallic cabinets containing the line modules, multiplexing unit, and a transmission and interface module (TxM/IM), which connects the cabinet to the CO, generally using fiber-optic cable. Sometimes DLC remote terminals are installed in vaults or on poles instead of in cabinets. While a copper wire connection between a subscriber and its CO cannot expand for more than 3.5 to 4 km, a fiber-optic cable

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

between a DLC remote terminal and a CO can reach lengths of 7 to 10 km. Examples of DLC remote terminals are shown in Fig. 2.58 and in other figures throughout this book. In modern wire-line networks, DLC systems are replaced by broadband cabinets that are connected to the CO with fiber-optic cables. Externally, these broadband cabinets look like DLC remote terminal cabinets and they are often installed along the curb of sidewalks or roads, but broadband cabinets are able to provide a wide variety of communication services in addition to the only communication service provided by traditional POTS networks, which was connectivity for voice calls. Remote DLC terminals and broadband cabinets are widely used around the world, and they represent edge network components.

Nowadays, broadband communications utilizing broadband cabinets to connect to the user devices have enabled voice calls that can be originated from a wide range of devices. In some cases, the devices are part of a packet-switched network, such as an Ethernet network (the preeminent Local Area Network), instead of the circuit-switched PSTN. These calls usually follow a set of protocols that are collectively designated as "voice over IP" (VoIP) technology, because, as they are packet-switched, they usually are routed through the Internet. It is certainly possible, and quite common, for a party in the call to be VoIP while another (or others if being a multi-party conference call) is a PSTN party or a cell phone. In these scenarios, it is necessary to interconnect the call across networks with differing underlying technologies. This is achieved through a device called a "gateway" that is tasked with converting protocols and technologies between two different networks. In the case of a PSTN, gateways can be found at a central office, and for a cellular network, gateways form part of the core network, as discussed later in this chapter. Also, VoIP calls can potentially be routed end to end through the Internet. In these cases, "calls" should be thought of in a broader way, since they may include video "calls" (either conversational or streamed) and regular data exchange (web browsing, file exchange, etc.) In fact, in these cases "calls" are called "sessions." Nevertheless, while the nature of the technology used to establish packet-switched sessions is very different from the circuit-switched technology used in the PSTN, the broad topology of the networks is quite similar. In a packet-switched network, sessions are first connected through a "router" to the larger network (in what would be parallel to a local office). From this point, the session would be routed through switches at different levels of network hierarchy, where higher levels of hierarchy can be thought of as corresponding to aggregating a larger volume of traffic. In this way, the equivalent to the PSTN's tandem offices is the Internet's backbone switches. It is because of these parallels that the electric power infrastructure for packet-switched networks is also similar to the one for the PSTN.

Cable TV (CATV) communication networks, which in their initial operational years broadcasted television signals through wired connections to the user's TV, evolved into providing a variety of broadband services, including VoIP, with bidirectional communications, namely, TV signals or data are transmitted from a main station to the subscribers and also data signals are transmitted from the subscribers to the station. Figure 2.59 shows the basic architecture and elements of a CATV network. The main transmission station is called the head-end (H/E). To support Internet-based

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

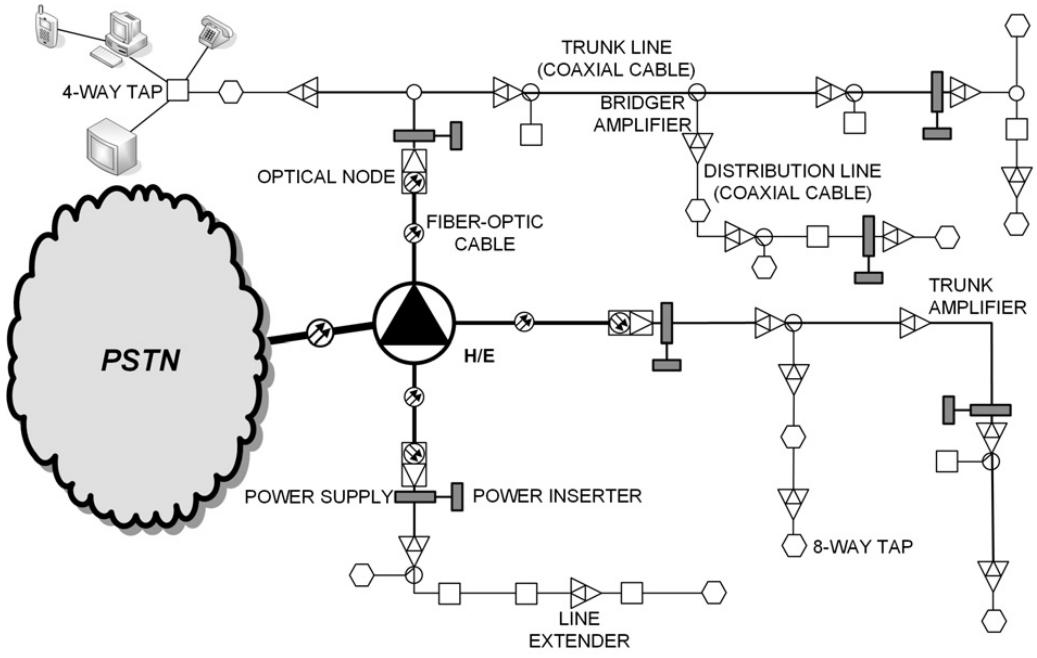

Figure 2.59 CATV network.

applications, the H/E is connected to the PSTN though at least one fiber-optic link. The H/E is linked with the subscribers with fiber-optic cables up to optical nodes and then with coaxial cables. In large networks, the H/E connects via fiber-optic cables to distribution hubs, each of which in turn connect to various optical nodes. Since the signal loses quality as it moves along the coaxial cable, it needs to be improved by using amplifiers or it is used without amplifiers for the last few meters from an optical node to the users. Depending on the size of the coaxial cable, the number of subscribers, and the topology, typically in cities the amplifiers are placed every few hundred meters.

The H/E, optical nodes and amplifiers require electrical energy to operate. Even though the H/E and the optical nodes require one power supply each, several amplifiers can be fed with one uninterruptible power supply (UPS). This is accomplished by injecting ac power into the coaxial cable, which is rectified at each amplifier. A UPS is a cascaded combination of a rectifier and an inverter with batteries connected at the rectifier output terminal to provide backup power for a few hours in case of an outage in the electrical supply. Usually, UPSs for individual amplifiers are placed in small cabinets mounted on poles, as shown in Fig. 2.60 – which may weaken the pole footing due to the extra weight of the UPS. Because these UPS only provide a few hours of operation from the batteries, it is necessary to use other power backup solutions during long power outages. The common alternative for providing backup power during long power outages is to use generators, but as explained in Chapter 7, this approach has some practical issues.

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

Figure 2.60 (a) A CATV UPS mounted on a pole and (b) a similar CATV UPS also mounted on a pole with its door open, showing the batteries on the bottom shelf.

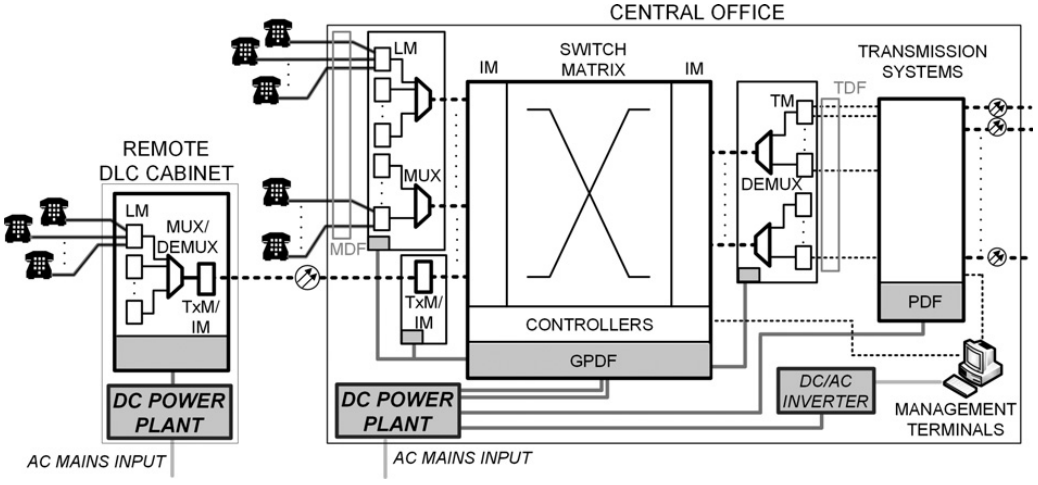

Figure 2.61 CO and DLC main components and their electrical supply scheme.

PSTN COs also contain all the ancillary services essential for the system to operate, among them, the direct current (dc) power plant. As shown in Fig. 2.61, the power plant receives alternating current (ac) electrical power and rectifies it into dc electrical power. It is distributed to a global power distribution frame (GPDF), the other system power distribution frame (PDF), and to inverters that feed the management terminals with ac power. The GPDF and PDF hold the fuses that protect the system in case of a short circuit in the equipment frames. The switch, transmission systems, and

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

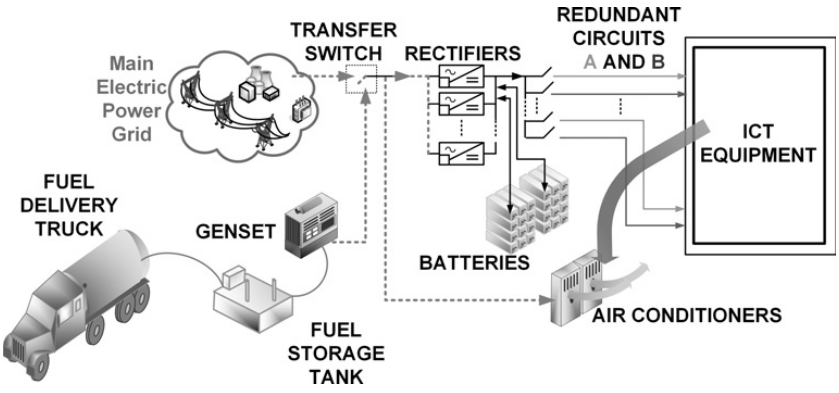

Figure 2.62 Basic elements of a telecom power plant and connection scheme.

Figure 2.63 Rectifier cabinets (right) and batteries (left) in a CO flooded by Hurricane Katrina.

management terminals are the main power plant loads. In traditional PSTNs in which connections from the CO to the subscribers are established exclusively using copper cables, the CO power plant also feeds subscriber telephones. Figure 2.61 also shows that the DLC remote terminal cabinet requires a separate power plant that receives the ac power from a local connection. A local power plant is the most commonly used and necessary approach to power both DLC remote terminals and broadband cabinets connected to the CO with fiber optics. However, as discussed in more detail in Chapter 7, there are various technologies used to power these edge network nodes.

One of the main functions of the power plant is to provide energy to the system even when there is an outage in the ac utility grid. This is accomplished by including batteries directly connected in parallel to the load and a combustion engine/electric generator set (genset) in a standby mode connected through a transfer switch to the ac mains input. Figure 2.62 shows a basic scheme of a telecom switch power plant, whereas Fig. 2.63 depicts a typical telecom power plant in the aftermath of a natural

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

disaster. During normal operation, the system is fed from the electric grid through rectifiers that convert the ac mains into dc power. The function of the batteries during an electric utility grid outage is to provide power to the system for a short time until the genset starts and the transfer switch connects the generator to the rectifier input. In this manner, as long as the genset has enough fuel and does not fail, the CO can operate normally until the electrical utility power is restored. The batteries are also engineered so that they can maintain the CO operating for a few hours in case the genset fails or the CO is not equipped with a permanent genset, as sometimes happens in small facilities. Extending battery capacity to more than a few hours is not practical due to their weight, size, and cost, and because without a generator, the air conditioning system, which has a power supply only backed up by the genset, will stop operating and the equipment within the CO will stop operating due to increased temperatures in the facility.

One problem with batteries is that they are very heavy because, usually, they are made with lead. Because they are heavy, the floor loading for a battery string may easily exceed 1 tn/m^2, which is the standard floor-loading for dwellings. This is a reason why in COs batteries are placed at ground level, where it is easier to reinforce the floor. However, this location makes them vulnerable to floods. Another problem with batteries is the difficulty in replacing them during periods of high demand, because manufacturers keep only a small battery inventory, as they need to be kept charged while stored. Thus, lead times for large orders may reach several weeks, so it is not uncommon to observe long replacement times after disruptive events in which a significant number of batteries are damaged.

One of the most important characteristics of the PSTN is its extremely high availability with downtimes that are expected to be less than a minute per year. This has both commercial and emergency (911 system) implications. Figure 2.64 shows a scheme of the enhanced 911 (E911) system in the United States with its three main elements: the PSTN, the public safety answering points (PSAPs), and the E911 offices. In the E911 system, the PSTN connects the call to the PSAP, as routed by the corresponding E911 office. Since CO areas do not generally coincide with PSAP areas, there is a database that indicates the E911 offices that route the calls to the corresponding PSAP center of the calling party. In Fig. 2.64, both party A and party B belong to CO X, located in PSAP area B. When party B calls 911, there is no issue because CO X, PSAP B, and calling party B are all located within the same PSAP area. However, when party A, located in PSAP area A, calls 911, the E911 office #1 routes the call through CO X to PSAP A. The system is also programmed to reroute 911 calls to backup E911 offices in case the primary one ceases to operate. For example, in Fig. 2.64, E911 office #2 is the E911 CO #1 backup.

In the PSTN not all the COs have the same importance. To reduce the cost in transmission links, the PSTN architecture has a radial configuration in which several COs are connected through another switch called a tandem or toll office. Hence, tandem switches are more important than regular switches, called class-5 CO or local offices (LO), that handle subscriber calls. Figure 2.65 presents an example of how tandem offices interconnect class 5-COs. The figure shows that party D can talk to party C only through tandem office Y. In the same way, party E can talk to party A only

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

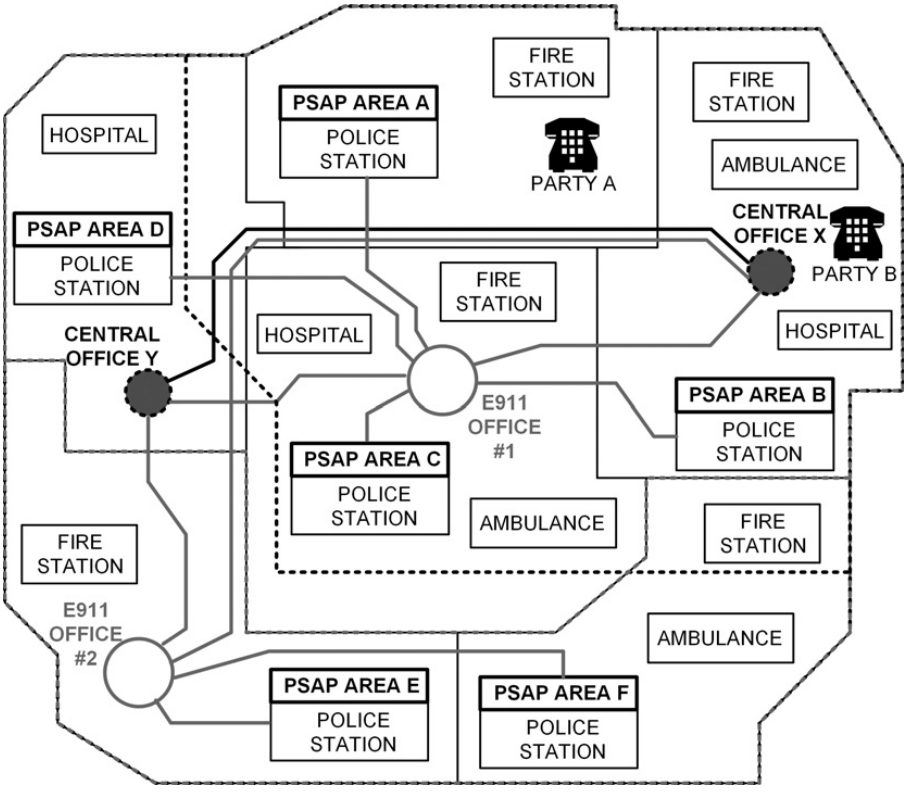

Figure 2.64 E911 system architecture representation.

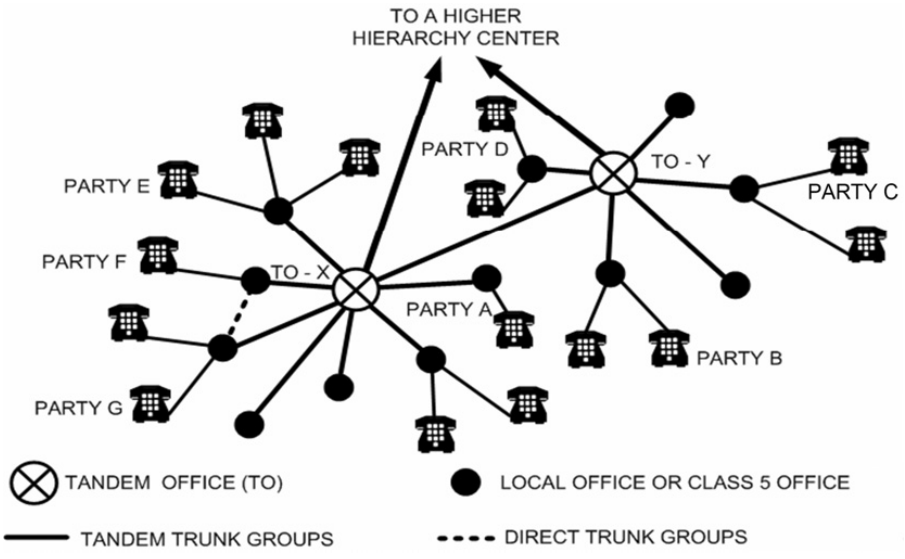

Figure 2.65 Typical CO connections in a city.

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

Figure 2.66 Microwave radio repeaters on top of a mountain.

through tandem office X. Moreover, party A and party B can only talk through tandem offices X and Y. If the traffic between two class-5 offices is high enough, they can be directly connected with high-usage trunks. For example, in Fig. 2.65 party F and party G can be connected without passing tandem office X. The importance of a tandem office becomes evident in Fig. 2.65. When tandem office Y fails, the subscribers of the five class-5 COs connected to it will only be able to talk to other subscribers of the same LO. Usually, tandem switches do not interconnect subscribers, but they connect class-5 switches to other, higher-hierarchy centers.

Nowadays, connections among COs are done with fiber-optic cables, which are even used for international links, sometimes even using transoceanic cables. Such cables are connected to large facilities, such as the one in Fig. 2.53, which provides access to 15 subsea cable landings. Fiber optic links require the use of repeaters every approximately 100 km to boost the signal. In transoceanic cables, these repeaters are powered with a conductor running at the center of the fiber-optic cable. Alternative, long-distance connections can be established with satellite links. Today it is still possible to find microwave radio systems for lower-capacity links. Sites exclusively used for microwave radio communications are thus smaller than those used for high-capacity connections. One issue with microwave radio connections is that they require direct line of sight between the antennas at both ends of the link, which limits the distance between immediate antennas to no more than 65 km with flat unobstructed terrain. For longer distances, repeaters, such as the one in Fig. 2.66, are used. Because of the need for line-of-sight communications, in mountainous regions microwave repeaters are placed on top of high mountains (e.g., see Fig. 2.67), which have difficult

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

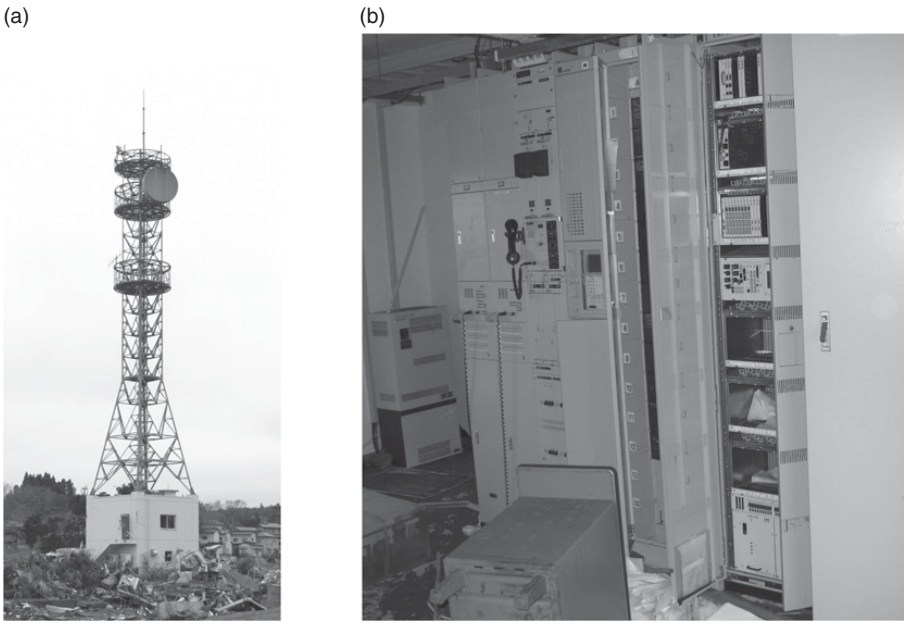

Figure 2.67 (a) Exterior and (b) interior of a microwave radio site in Saichi, Japan, after the 2011 earthquake and tsunami. The destroyed power plant was at ground level.

access and, in many cases, require autonomous power solutions using diesel generators that are refueled by helicopter or, if possible, a vehicle and which sometimes are supported by photovoltaic panels to reduce the generator refueling frequency.

The PSTN components within a CO area are divided into outside-plant and inside-plant components. All the elements located outside the CO up to the vertical blocks of the MDF are part of the outside plant. The remaining elements situated inside the CO are part of the inside plant. For historical reasons, sometimes transmission fiber-optic cables and microwave antennas are considered part of neither the inside nor the outside plant. The basic terminology used for outside-plant hardware is shown in Fig. 2.68: poles, manholes, a cable entrance facility, serving area interfaces, line drops, feeders, and distribution cables. In POTS networks those cables were formed of multiple copper conductors. More modern networks use fiber-optic cables from broadband cabinets replacing DLC remote terminals to the user devices at their homes or businesses. Although Fig. 2.68 shows typical overhead outside-plant components, in many urban and some suburban areas, outside-plant cables are installed buried and only connection boxes and tap enclosures are left to be seen aboveground, as exemplified in Fig. 2.69.

2.2.4 Wireless Networks

Wireless networks deserve an extra subsection to discuss the different architectures in which they may be set up. These two architectures are the *infrastructure type* of network and the *ad hoc type* of network.

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

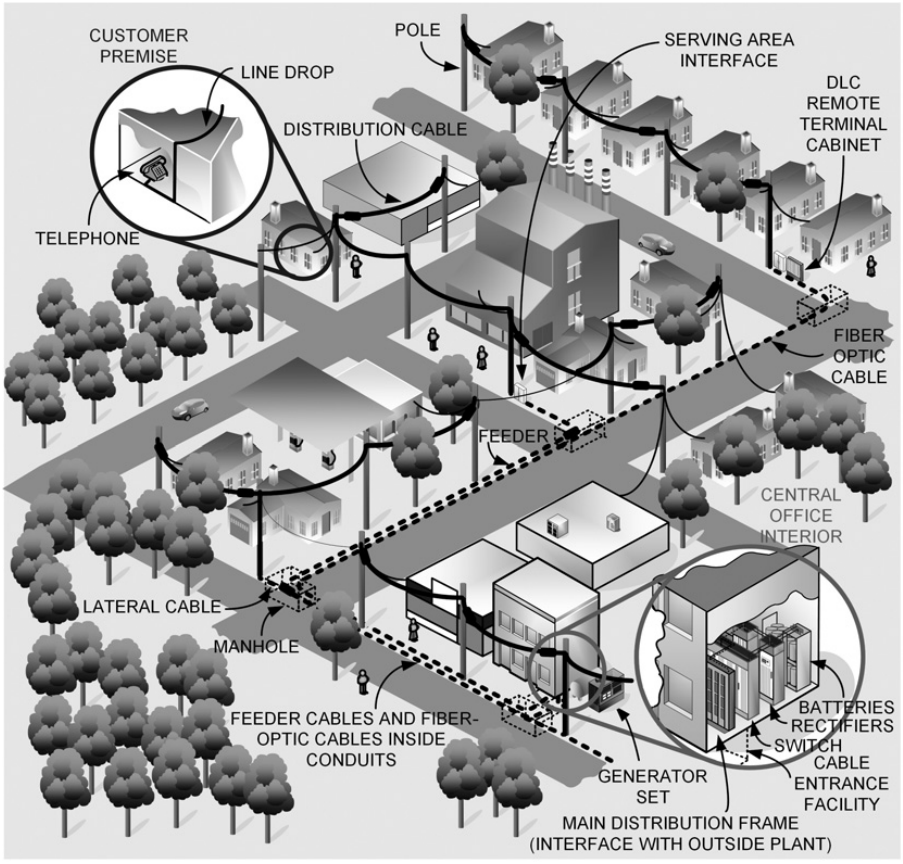

Figure 2.68 Main outside-plant elements of a CO area.

An infrastructure-type wireless network is composed of an access point and wireless terminals. The access point also receives the name of a base station, and a wireless terminal may also be called a user equipment. In an infrastructure-type wireless network, the wireless terminals communicate with each other and to outside their own network (e.g., to the Internet) by connecting through the access point. This is a centralized architecture where the wireless terminals establish links only with the access point. The access point performs not only the task of providing connectivity to the wireless terminals, but it also controls when data is sent to the wireless terminals, when terminals can transmit (to avoid transmission "collisions" between multiple terminals), and even many transmission parameters for the wireless terminals (one such example could be the transmit power). Because each terminal communicates through a wireless link to the access point, it is considered that in an infrastructure-type network the communication is through a single wireless "hop" (a single point-to-point connection). This is considered in this way even when a terminal is communicating with another terminal that is connected with the same access point because this type of communication usually

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

Figure 2.69 Underground telephone outside plant network with tap enclosures and connectors marked with circles.

is instantiated as two Data Link layer point-to-point connections and does not involve network routing.

In contrast, in an ad hoc network all nodes are of the same type. That is, there is not a specific node that centralizes the communication between other nodes or to other networks, as is the case with the access point in an infrastructure-type network. In an ad hoc network, each node has the capability to join the network by identifying and establishing links with other neighboring nodes. Because of this, ad hoc networks have a self-organizing property and operate in a decentralized way. Moreover, it is common in ad hoc networks that nodes communicate by establishing routes consisting of multiple wireless "hops" (multiple point-to-point wireless links) through the network. However, by their own nature, ad hoc networks are more difficult to manage and it is quite challenging to consistently meet performance guarantees on the different ongoing instances of communications existing in the network. Because of this, the most common types of wireless networks used by the general public are of the infrastructure type.

Cellular networks are perhaps the most common form of wireless network used by the general public. Cellular networks have been designed to make best use of the limited resource that is the radio spectrum so that they can accommodate as many simultaneous connections as possible. Cellular networks operate by fitting wireless

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

transmissions in a designated portion of the radio spectrum, called a radio spectrum band. Today, a typical radio band for a cellular network may have a bandwidth of 20 MHz at a central frequency that is between roughly 1.8 GHz and 3.7 GHz (these numbers vary from country to country). In order to fit multiple simultaneous connections, the spectrum band is subdivided into portions associated with different frequencies that receive different names depending on the cellular network technology. (They may be called "channels," "subchannels," "resource elements," etc.) However, this subdivision of the radio spectrum is still not sufficient to accommodate the large number of high-data-rate connections that a cellular network needs to service. To achieve this, the architecture of cellular networks has been designed to reuse the same radio frequencies while controlling the interference from other same-frequency transmissions to within acceptable levels. This design constitutes the defining characteristics of cellular networks and, even more, is the reason for the name "cellular" given to these networks. The central elements of the cellular network architecture are as follows:

1) Following an infrastructure-type architecture, end users connect through a wireless link to an access point, generally called a *base station*. The wireless device that users employ to connect to the base station is called *user equipment* (UE). The collection of connections between UE and base stations is called the *Radio Access Network* (RAN). The base stations are further connected to the core network, which is the network linking the base stations to nodes managing the operation of the overall network and also connecting to other networks (including the Internet).
2) The complete geographical area that is serviced by a cellular network is divided into smaller zones, each associated with the coverage area of one base station. The coverage area of a base station is controlled by adjusting its transmit power. By controlling the base station's transmit power and implementing different interference control/mitigation techniques, base stations that are nearby (and even neighboring base stations) can transmit using the same radio frequencies, effectively reusing the radio spectrum bands for different users located in different locations of a cellular network service area. With this approach, the number of connections that could be supported over a radio spectrum band is multiplied by the frequencies' reuse factor. The smaller the coverage area of base stations, the larger the density of base stations is, the larger the reuse factor becomes, and the more simultaneous connections that can be supported by radio spectrum band. Because of this, a cellular network may be comprised of hundreds of base stations, especially in cases where a cellular network needs to support a very large number of simultaneous connections (for example, in a large city).

Figure 2.70 illustrates different elements that are present in a typical architecture for a cellular network. In the base station, the signal processing is divided into a baseband unit and a radio unit. The baseband unit performs the processing for signals with a spectrum with a central frequency equal to zero. The radio unit performs the processing for signals with a spectrum with a central frequency that is in the radio spectrum band assigned to the cellular network. Traditionally, the radio and baseband

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

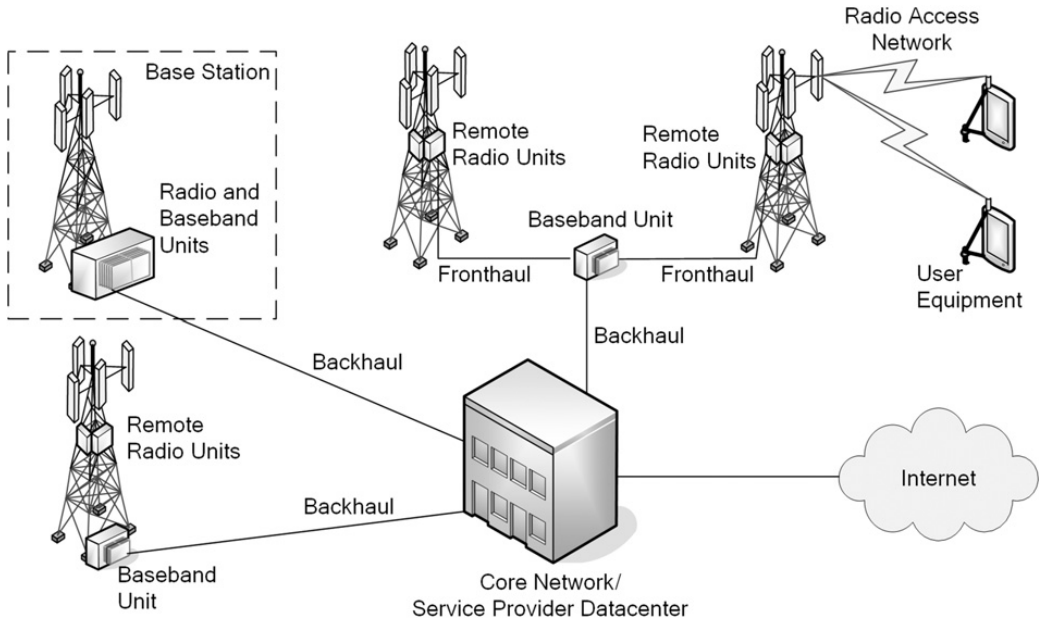

Figure 2.70 Cellular network architecture.

units are collocated in a rack inside a cabinet next to the base station antenna tower. However, advances in integrated circuits have allowed engineers to separate the two units and move the radio units closer to the antennas in what is known as a remote radio unit. Placing the radio units closer to the antennas presents the advantage of reducing the signal power loss over the cable connecting the radio units to the antennas. Also, advances in computing capabilities have allowed engineers to implement baseband units that do the signal processing for multiple base stations. Today, these shared baseband units can be seen decoupled from the base station and implemented in virtualized environments at a centralized location. This centralized implementation can be a cloud server (in what is called a cloud RAN, C-RAN). When the radio and the baseband units are separated, the data connection between the two is called the *fronthaul*. Similarly, the connection between the baseband unit (be it at a base station or in a cloud server) and the core network is called the *backhaul*. Backhaul connections are usually established with fiber-optic cables or microwave links and less commonly with a satellite link. Wireless networks have a diverse topology, particularly among core network elements, which implies that there is more than one path linking two nodes. Finally, it is worth noticing that at the point where base stations are connected to the core network there is usually a connection also to the service provider datacenter. This connection to the datacenter is used to manage the handoff of connections between base stations when necessary due to UE mobility and to implement a number of functions associated with the validation of the identity of a UE. Connections handoff requires establishing a database that is dynamically updated so the system can keep track of the UE locations to allow connections to be routed to the

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

appropriate base station. An important step after a network outage is to restore this database. However, because the UE location may have changed during the outage, the database changes, too, which makes the database restoration process sometimes time consuming and, thus, delays bringing the network back into operation even after all physical components are back in operation. Cellular networks are not only connected to a datacenter to manage handoffs, but they may also connect to datacenters in order to provide data services. Additionally, cellular networks are connected to PSTNs in order to enable voice calls with PSTN users, to connect to other parts of the wireless networks using data transmission assets of the PSTN, to provide emergency call services (i.e., allow connecting to a 911 PSAP), or to enable data services by connecting to the Internet. These connection points to the PSTN are called channel service units (CSU). Moreover, wireless network equipment may be collocated at PSTN facilities when both networks are subsidiaries of a same company, such as AT&T and AT&T Wireless in the United States, Bell Canada and Bell Mobility in Canada, BT and BT Mobile in the United Kingdom, Deutsche Telekom and T-Mobile in Germany, Telefonica and Movistar in Spain, or NTT and NTT-Dokomo in Japan, to cite a few examples from around the world.

The electrical power supply of the core network and base stations is similar to that of the PSTN. As Fig. 2.71 represents, the core network is analogous to the CO in the PSTN, and base stations play a role analogous to the one DLCs play in PSTN and edge network nodes. However, the need for a wired connection in the case of the PSTN makes this network more susceptible to damage during natural disasters than cellular networks. Yet, because wired connections allow providing power to users' telephones,

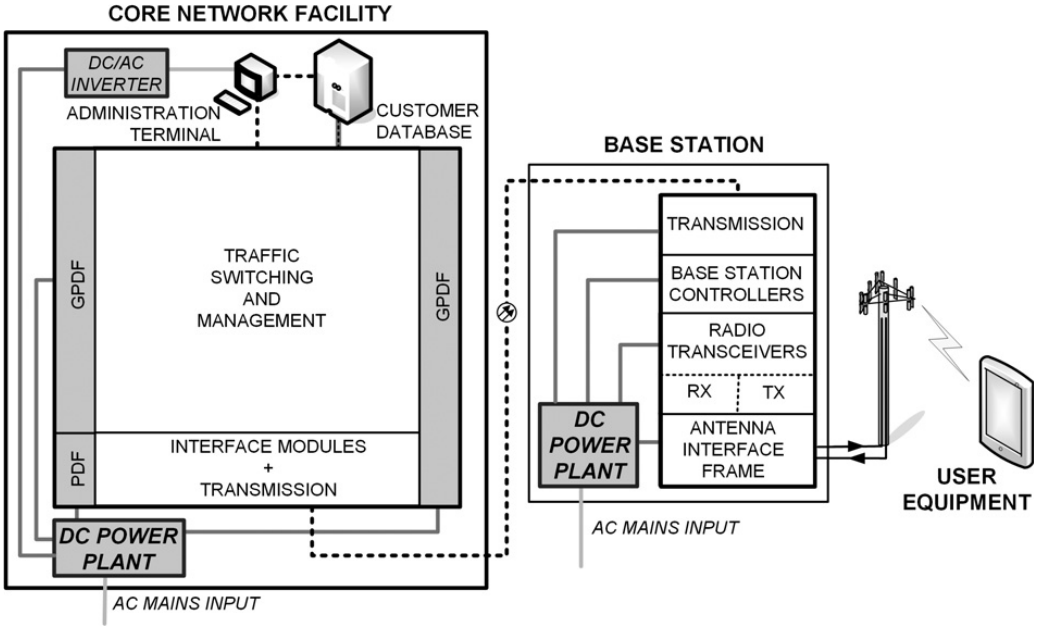

Figure 2.71 Cellular network main components.

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

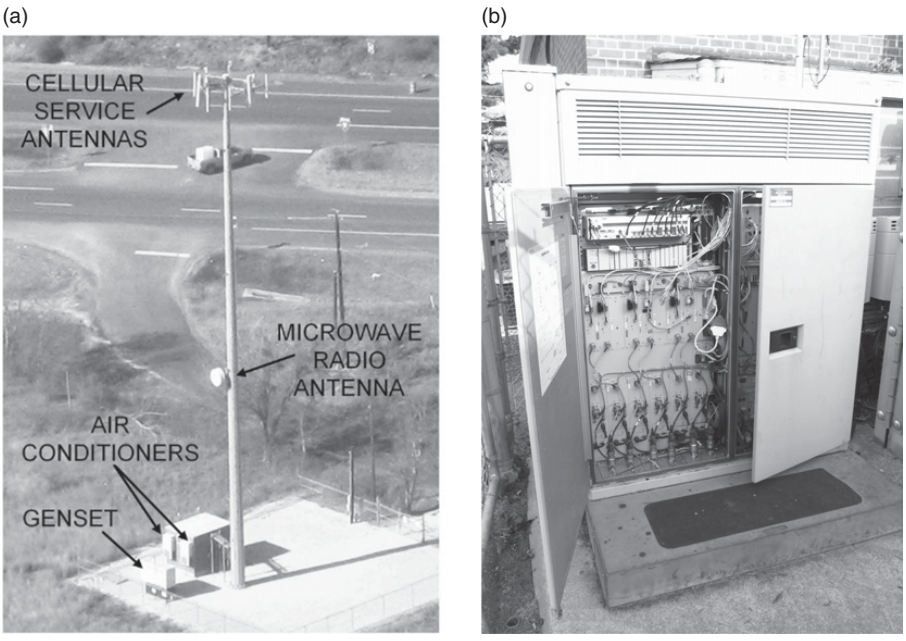

Figure 2.72 (a) A typical macrocell and (b) a smaller macrocell, showing the interior of the electronics cabinet.

PSTN tends to be less susceptible to disruptions caused by electric power grid outages than cellular networks because, as shown in Fig. 2.71, UEs and base stations are no longer fed directly from a core network element. For this reason, UEs need to have their own batteries that must be recharged regularly, thus creating resilience issues during long electric power grid outages that tend to follow natural disasters and other disruptive events. Base stations also have the same general power plant architecture, although inclusion of a permanent generator set depends on the operational practices of each network operator and the type of base station with respect to their coverage range. For example, in the United States "large" macrocells, such as the one in Fig. 2.72, have been equipped over the years with permanent gensets to improve resilience even during moderate disruptive events because most macrocells require air conditioners whose power supply is not backed up by batteries so as to cool not only their electronic equipment but also their batteries to prevent loss of life due to high operating temperatures. "Small" macrocells, such as the one in Fig. 2.72, in the United States have also been equipped with a permanent genset over the years even when they may be less commonly equipped with air conditioning systems. Instead, these base stations may have fans powered directly from the dc battery bus to circulate the air, bringing air at ambient temperature in and pushing hotter air out. Nevertheless, use of permanent gensets for all types of macrocells is less common in most other countries around the world. Also, use of permanent gensets for small base stations, such as those in Fig. 2.73, is very uncommon around the world because of the space limitations for a genset and practical issues for their fueling due, in part, to the increasing number of

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

Figure 2.73 Examples of small base stations; from (a) to (d), a cell site in downtown Moscow (the building in the background is Russian State Library), one in Kuala Lumpur near the National Mosque of Malaysia, one in downtown Christchurch, New Zealand, and one in western Pennsylvania.

them being deployed. Moreover, because batteries for extended backup times tend to occupy a relatively large volume and are heavy, batteries' backup times for small base stations mounted on poles tend to be relatively short even when these small base stations consume less power than macrocells. As a result, small base stations have limited electric power backup autonomy, which tends to decrease resilience, as further discussed in Chapter 7, which also includes an explanation of power supply alternatives for edge nodes in communication networks. Still, limited backup time in small base stations creates resilience and availability issues in the latest wireless network technologies, such as 5G, due to the increased use of micro-, nano-, pico-, and femtocell sites in these networks.

The most distinctive infrastructure feature of base stations is, arguably, their tower for the antennas. Macrocells have two types of towers, exemplified in Fig. 2.74: metallic monopoles and lattice towers. Sometimes, base stations use towers of high-voltage electric power transmission lines, as shown in Fig. 2.75. Although this design reduces costs for both electric power utilities and wireless network operators, they also create maintenance and servicing issues due to the presence of high voltages in the nearby conductors. On other occasions, antennas are mounted on water tanks or industrial stacks, as exemplified in Fig. 2.76. In many instances, especially in the United States, towers are owned by a third-party company, which leases part of the tower to different network operators that share the same tower. One common complaint about these towers is their negative aesthetic impact. Hence, sometimes towers are disguised as

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

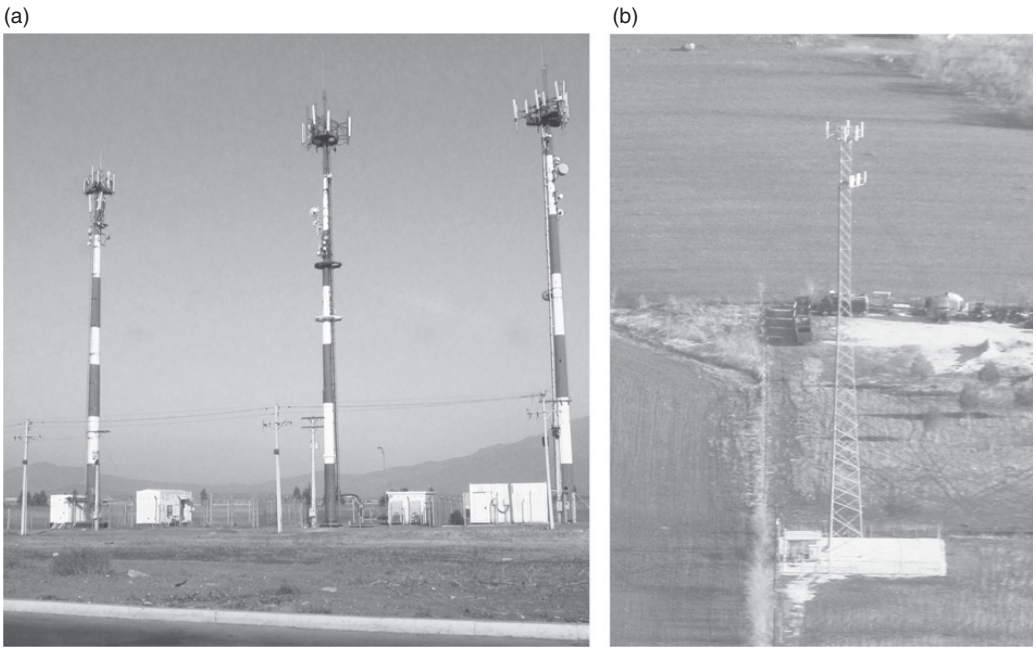

Figure 2.74 (a) Three metallic monopole towers and (b) one lattice tower.

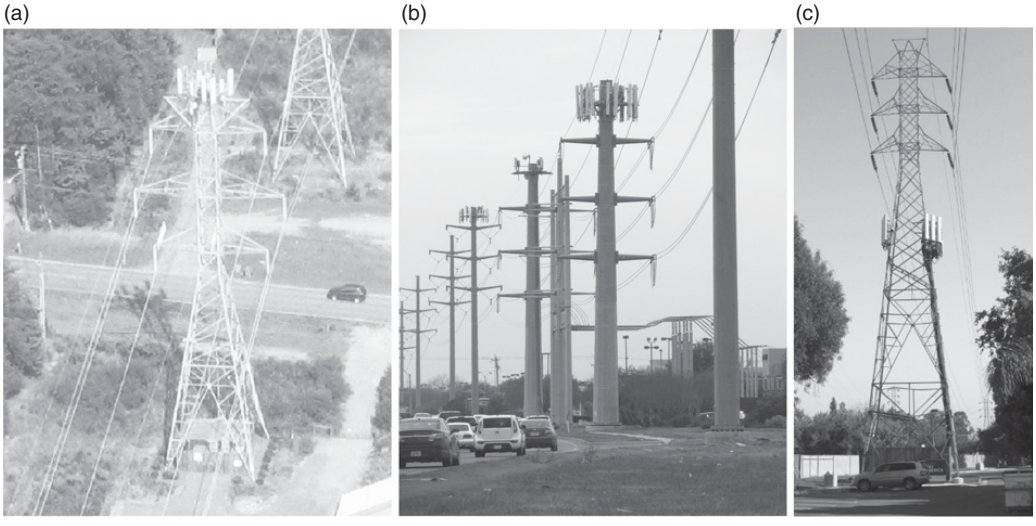

Figure 2.75 Examples of cellular antennas mounted on high-voltage transmission lines. The antennas in (c) are mounted lower, perhaps to avoid issues with the high-voltage conductors when servicing the antennas.

trees, as shown in Fig. 2.77. Another way of reducing this aesthetic impact applicable to urban environments may be to place the antennas on building rooftops, as exemplified in Fig. 2.78, although many rooftop-mounted cell sites still may include noticeable towers, as shown in Fig. 2.79. However, this solution creates accessibility and structural

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

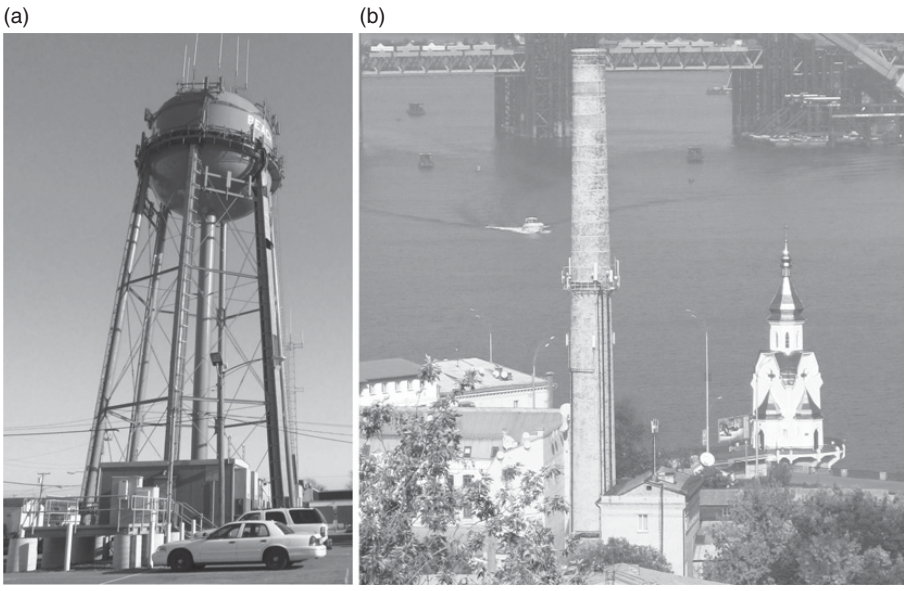

Figure 2.76 Cellular antennas mounted on a water tank (a) and on a stack (b). Another cellular antenna, mounted on a rooftop, is seen in the image on the right.

Figure 2.77 Examples of cellular towers disguised as trees.

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

Figure 2.78 Two examples of rooftop-mounted cellular antennas.

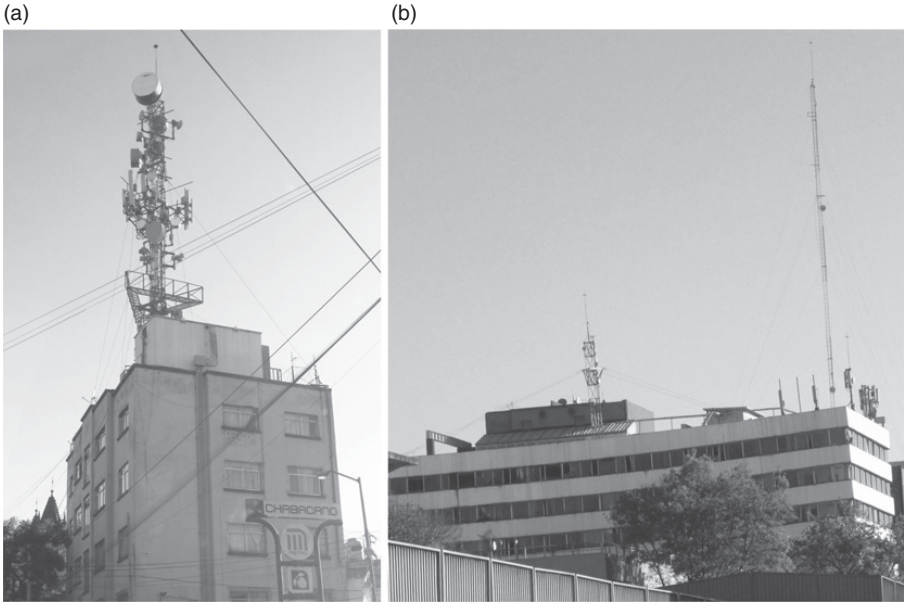

Figure 2.79 Examples of cellular towers on building rooftops. The building in (b) partially collapsed during the 2017 earthquake in Mexico.

issues that are also discussed in Chapter 7, because these aspects are dependent on the building operations and construction. For example, it may not be possible to have an air conditioning system for the base station independent of that used for the building without affecting the building aesthetic by installing separate external air conditioners,

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

Figure 2.80 Rooftop cell sites with (a) added air conditioners and (b) shelter.

as shown in Fig. 2.80, or by placing the base station inside a shelter, also as exemplified in Fig. 2.80. Small cell sites, such as the microcells in Fig. 2.81, usually have simpler towers, typically relying on conventional wooden, metallic, or concrete poles, which are sometimes shared with other infrastructures, such as those for street lighting.

It is worth noting that satellite networks have a structure similar to that of regular mobile wireless networks. Evidently, a main difference is that the base stations are replaced by satellites orbiting the earth and powered by solar energy and batteries. Core network elements are still located on land and are powered by the electric utility.

2.2.5 Other Information and Communication Networks

It is also relevant within the context of natural and man-made disasters to highlight the importance of traditional TV and public radio systems, which broadcast signals from a main station using antennas. These antennas are vulnerable to high winds. However, if they fall, the transmission can easily be shifted to some other location within the broadcasting area with an appropriate antenna. Traditional radio and TV systems are useful during a disaster aftermath to broadcast messages of public interest, such as information about the location of food and water distribution centers, or for immediate impending disasters in order to alert people about a potentially dangerous situation, such as, for example, earthquake warnings in Japan or tornado warnings in the United States. Their signals are emitted from antennas typically located in very tall lattice towers, such as that in Fig. 2.82, or, in the case of city centers, on top of skyscrapers or

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

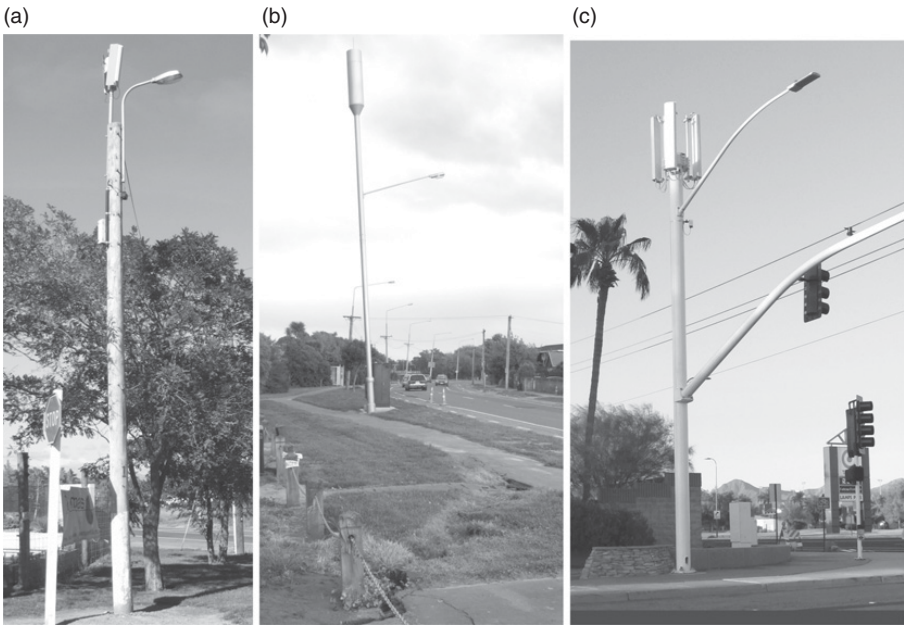

Figure 2.81 Examples of small cell sites mounted on streetlight poles. The earth crack shown in (b) was caused by an earthquake.

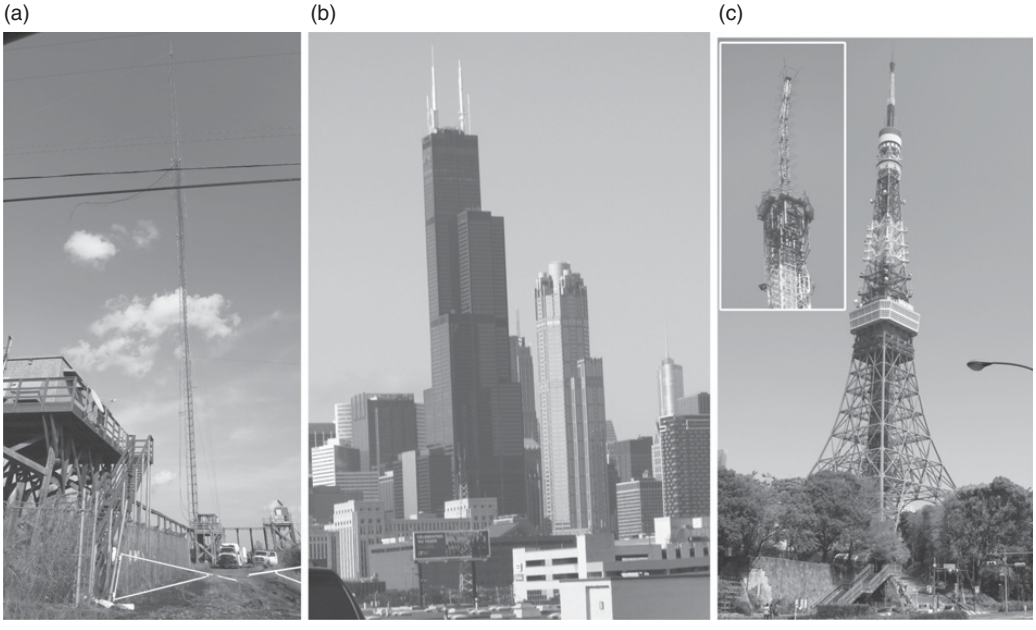

Figure 2.82 Examples of broadcast radio and TV towers and antennas: (a) a broadcasting radio antenna damaged during Hurricane Isaac when the cable connection was severed, (b) antennas on top of the former Sears Tower, and (c) the Tokyo Tower showing the damage to the antennas on top of the tower from intense shaking during the 2011 earthquake in Japan.

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

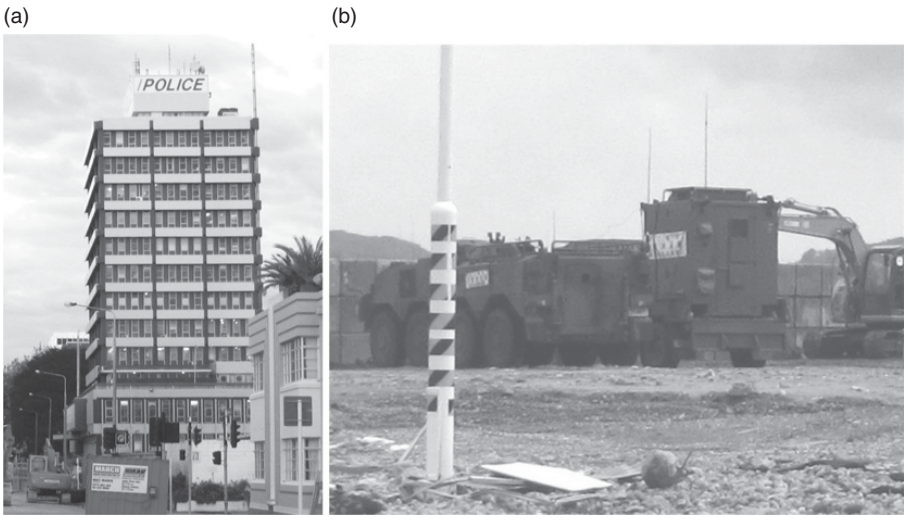

Figure 2.83 (a) Christchurch's police building with communication antennas on its roof after the 2011 earthquake and (b) military communications equipment in Osawa, Japan, after the 2011 earthquake and tsunami.

purposely made towers, as also exemplified in Fig. 2.82. Two-way radio systems, such as those used by police and other emergency services, are also useful, but their use is limited to operators in the security and emergency response forces. One desired characteristic of these systems is interoperability to make it simple to coordinate activities among different agencies and jurisdictions. Their towers are typically located in buildings belonging to safety and security services (e.g., see Fig. 2.83) or emergency response offices. In many disasters these networks are also supported by armed forces communications equipment, such as the one in Fig. 2.83. In many countries, amateur radio (also known as ham radio) systems are commonly used during disasters. In the United States it is relatively common to have a few of these radio terminals in emergency management offices as a last resource for communicating to the public.

2.3 Queuing Theory Overview

As indicated earlier, a packet-switched network is also called a store-and-forward network because it is based on the idea that packets, when they arrive at a network node, can be placed in a queue where they can wait until the node can process and route them to their next hop toward some final destination. Because of the central role of waiting queues in this operation, the functioning of network nodes and the process that a packet undergoes at a store-and-forward network node from its reception to its transmission, is studied based on the mathematical framework of queuing theory. The central element of study in queuing theory is the queuing system. Figure 2.84 shows the main elements composing the simplest queuing system, known as a single

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

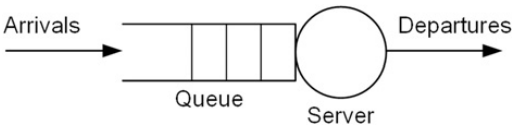

Figure 2.84 Packets over in the protocol stack.

server system. A queuing system is formed by a queue, which is a memory buffer where the received packets wait to be processed, and a server, which is the computing element tasked with processing a packet and sending it onwards through the next network node in the path to the final destination. The characteristics of the packet traffic that arrives to the queue is described through an *arrival process* or *input process*, which is most often a random process. How packets are placed and removed from the queue is described through a *queue discipline*. The most common queue discipline is *first-in-first-out* (FIFO), which, as the name indicates, describes a queue where packets are passed on to the server in the order they arrived at the queue. This queue discipline is also known in other contexts as "first come, first served." Other queue disciplines are *last-in-first-out* (LIFO), *random service order* (RSO), and many others considered for more specific cases. When the packets are removed from the queue, they are passed on to the server (or, in some cases, one of multiple servers that work in parallel to each other to service the queue). The operation of the server is characterized in terms of the time, called the *service time*, that it takes for a packet to be serviced and transmitted out of the server and, consequently, out of the queuing system. It is often the case that the time to process a packet is negligible compared to other times in the queuing system, in which case the service time amounts to the time it takes to transmit the packet. Since transmission from the server occurs at a constant transmit rate (a constant amount of bits per second), the service time for each packet is directly proportional to the size of the packet. Also, it is common to characterize the server by the inverse of the service time, called the *service rate* (which is, of course, inversely proportional to the packet size).

A widely established convention indicates the use of the *Kendall notation* to describe the configuration of a queuing system. The Kendall notation specifies the queuing system configuration through a triple written as *A/B/s*, where *A* represents the type of arrival process, *B* represents the type (usually of a random process) of the service time, and *s* is the number of servers removing packets from the queue. Some of the most common and also most interesting arrival processes are (along with the corresponding Kendall notation symbol): *M* for a Poisson process (characterized by exponentially distributed packet interarrival times), *D* for deterministic, fixed, or periodic packet interarrival times, E_k for an Erlang distribution made from the sum of k independent and identically distributed (i.i.d.) exponential random variables, and *G* for a general distribution (a distribution of independent packet arrivals for which no specific characterization is known or assumed). In terms of the nomenclature *B* that represents the service time, perhaps the most common and also most interesting arrival process is the one with an exponential service time, which arises because the packets

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

are transmitted by the server at a fixed transmit bit rate and have a size that is random following the exponential distribution, which is given by

$$f_X(x) = \mu e^{-\mu x}, \quad x \geq 0 \tag{2.1}$$

where X is the random variable (service time in this case) and μ is a parameter for the distribution that soon will be introduced as the *mean service rate*. The Kendall notation symbol for the exponential service time is M.

An interesting extension of the exponential service time is the case when the packet's service consists of a tandem of multiple stages of service, each of them following an exponential service time. If we consider that the total service time is comprised of a tandem of k stages, each following an exponential distribution with parameter $k\mu$, the resulting distribution for the whole service time can be shown to be

$$f_X(x) = e^{-k\mu x} \frac{(k\mu x)^{k-1} k\mu}{(k-1)!}, \quad x \geq 0. \tag{2.2}$$

This resulting distribution is called the Erlang distribution, and its Kendall notation symbol is E_k. Besides service times that follow the exponential or the Erlang distribution, other common cases are the deterministic, fixed, or periodic service time cases (because packet size is deterministic or fixed), which is represented by the Kendall notation symbol D, and a general distribution, which is indicated with the Kendall notation symbol G.

As an example of Kendall notation, the most studied and simplest queuing system is the one with one server, Poisson arrival process, and exponentially distributed service time, which results in the notation: *M/M/1* (if there were s servers, the notation would be *M/M/s*).

Telecommunication engineers make use of queuing theory to design networks and to predict the behavior of networks under different scenarios. In these works, engineers are commonly interested in modeling and calculating different operating variables that are indicative of the system's performance and/or its operational limits. Typical examples of these magnitudes are the average time a packet may experience in a queuing system or the average size of the queue (number of packets that it is holding) that is associated with an average delay, or the probability that a new call will not be granted access to a communication system because of a lack of available capacity. Some of these magnitudes can be calculated using simple relationships that are applicable to any queueing system. One of these useful relationships is Little's theorem, which indicates that $q = \lambda T_q$, where q is the mean number of packets in the system (those being served plus those waiting in the queue), T_q is the mean time that a packet spends in the system, and λ is the average number of packets that arrive to the system per unit time (usually measured in the unit of packets/second). Another form of Little's theorem concerns only those packets waiting in the queue by expressing that $w = \lambda T_w$, where w is the mean number of packets waiting in the queue and T_w is the mean time that a packet spends waiting in the queue.

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

One important parameter for a queuing system is the offered load, traffic intensity, or offered traffic, which is defined as the ratio of the total mean arrival rate to the single server mean service rate. The mean service rate, usually denoted as μ, is the inverse of the mean time to service a packet $\mu = 1/T_s$. Then, the offered load ρ is defined as $\rho = \lambda/\mu$ and, because it is a dimensionless quantity, it is measured in "Erlangs." One reason why the offered load is an important parameter is because it determines the validity range for the formulas modeling queueing systems. Specifically, all queuing models are valid for $\rho \in [0,1]$. When $\rho = 1$, the mean time that a packet spends in the system is infinity. Network traffic engineers tend to design networks so that the offered load does not exceed values of around 0.8 because, when ρ exceeds these values, the mean time that a packet spends in the system starts to increase very rapidly (often at a rate of $1/(1-\rho)$.

Next, after accepting the naturally intuitive idea that the mean time that a packet spends in the system is equal to the mean time that a packet spends waiting in the queue plus the mean time to service a packet, $T_q = T_w + T_s$, we can multiply both sides of this expression and use Little's theorem to find that $q = w + \rho$. This expression is for the case of a single-server system. In the case of systems with N servers, because the load gets distributed over all the servers, the formula for offered load is slightly modified as $\rho = \lambda/(N\mu)$ and, as a result, the expression for mean number of packets in the system becomes $q = w + N\rho$.

The simplicity in the expressions we have just discussed is that they do not require specific assumptions on the arrival traffic or service time statistics. For the modeling of other variables measuring the performance of a queuing system, usually a more specific analysis becomes necessary. Because queuing systems simply are a memory system, and because memory systems are usually modeled as being in different states (the states being defined as the number of packets in the queuing system), the analysis of queuing systems often revolves around the use of Markov chain random processes. One of the simplest, but also most applicable, systems is the *M/M/1* queuing system. For this system, and using Markov chain random processes analysis, it is possible to derive expressions such as the following:

- Mean number of packets in the system: $q = \dfrac{\rho}{1-\rho}$
- Mean number of packets waiting in the queue: $w = \dfrac{\rho^2}{1-\rho}$
- Mean time that a packet spends in the system: $T_q = \dfrac{T_S}{1-\rho}$
- Mean time that a packet spends waiting in the queue: $T_w = \dfrac{\rho T_S}{1-\rho}$
- Probability mass function for the number of packets in the system Q: $Q : P[Q = n] = (1-\rho)\rho^n$.

The natural extension of the *M/M/1* queuing system is the *M/M/s* queuing system where now the system has s servers. The study of this queuing system is also based on a Markov chain process with states defined as the number of packets in the system.

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

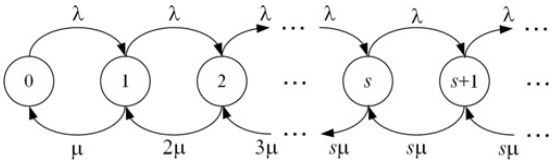

Figure 2.85 Markov chain modeling an *M/M/s* queue.

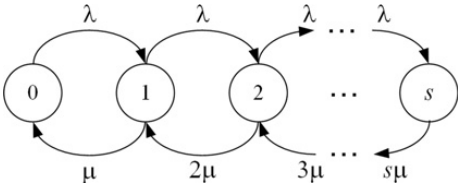

Figure 2.86 Markov chain modeling an *M/M/s/s* queue.

Figure 2.85 shows the diagram of the Markov chain modeling the *M/M/s* queue. As can be seen, the model assumes changes in the number of packets by one unit at a time. The figure also shows the rate of transitions between states. It can be seen that the arrival rate remains equal to λ for all states and, because servers process incoming packets in parallel, effectively distributing the workload among servers, the service rate for state I is equal to $i\mu$.

Nevertheless, an *M/M/1* queuing system is unrealistic in that it assumes that the queue has infinite memory. The consequence of this assumption is that it is not possible to model important events, as, for example, the case when an incoming call cannot be accepted into the system because the queue is full (or, equivalently, the communication system has no capacity). An important variation of the *M/M/1* queuing system is the Erlang loss model, which consists of the case where the queuing system has s servers. Figure 2.86 shows the diagram of the Markov chain modeling the *M/M/s/s* queue. As can be seen, the diagram is similar to the one shown in Fig. 2.85 for the *M/M/s* queuing system, only that now all the states beyond state s do not exist, which reflects the fact that these states are the ones where packets are placed in memory waiting for a server to become available. Because there is no memory in which to place incoming packets when all servers are busy, in an *M/M/s* queuing system, new packets that arrive under these conditions are simply denied access to the system (or, equivalently, are eliminated from the network). Because of this, it is important for engineers to know the probability of new packets arriving when all servers are busy, what is known as the *blocking probability*, as a function of the offered load $\rho = \lambda/\mu$. Following queuing theory analysis, it can be shown that the blocking probability p_b is

$$p_b = \frac{\frac{\rho^s}{s!}}{1 + \rho + \frac{\rho^2}{2!} + \ldots + \frac{\rho^s}{s!}}. \tag{2.3}$$

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

From this expression, the expected number of packets that would be blocked from entering the system per unit time can be calculated by multiplying the blocking probability by the average packet arrival rate: λp_b. With analogous thinking, traffic engineers also calculate the *carried load*, which is given by $\rho(1 - p_b)$.

2.4 Reliability and Availability Concepts

Reliability $R(t)$ of this entity is defined as the probability that an item or component will operate under specified conditions without failure from some initial time $t = 0$ when it is placed into operation until a time t. The definition of failure of a component can take different forms. For some components, such as a resistor or a capacitor or most other passive circuit components or semiconductor devices, a failure implies that it cannot operate meeting its intended function. For example, a capacitor experiences a failure when it can no longer store electrical energy according to its given capacitance. For other components, such as batteries, a failure occurs when it can no longer meet some performance requirements. For example, a battery can be considered to have failed when, at a given nominal temperature, its capacity falls below a given percentage of its nominal capacity. That is, for these types of components, some level of performance degradation is accepted without implying a failure condition. Notice that one key aspect of the definition of reliability is that it is defined as a probability. Hence, it can only take values between 0 and 1. Another key aspect of this definition is that the entity needs to operate without failure during the entire period of time under evaluation. That is, the repairing concept is implicitly not considered as part of the evaluation of component reliability. The complementary concept to that of reliability is called unreliability, $F(t)$. Hence, in a mathematical form it is

$$F = 1 - R. \tag{2.4}$$

That is, unreliability is the probability that an item fails to work continuously over a stated time interval. The explicit mention in this definition that the item needs to work continuously is related to the notion that the item should not experience any failure, as was mentioned in the definition of reliability. As a result of these notions, it is implicitly considered that the concept of reliability cannot be applied directly to repairable components or systems.

Reliability calculation requires one to define a hazard function, $h(t)$, first. A hazard function indicates the anticipated number of failures of a given item during a specified time period. That is, the unit of measurement for $h(t)$ is 1/hour, 1/year, or any other equivalent unit. For electronic and electrical components, the hazard function during the useful life period of electronic components is approximately constant. This constant value for $h(t)$ is conventionally named as the constant failure rate λ. With a constant failure rate, unreliability of a component is a cumulative distribution function that equals

$$F(t) = 1 - e^{-\lambda t}, \tag{2.5}$$

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

where the time is the random variable. Hence, the corresponding probability density function is

$$f(t) = \lambda e^{-\lambda t} \tag{2.6}$$

and reliability equals

$$R(t) = e^{-\lambda t}. \tag{2.7}$$

Thus, the reliability of an item with a constant failure rate is represented by an exponentially decaying function in which at time $t = 0$ there are no chances of observing a failure and in which there is almost a 37 percent chance of not observing a failure in such a component from the time it was put into operation until the time given by $1/\lambda$. The value of $1/\lambda$ has another very important meaning in reliability theory. Consider (2.9); the expected value for such a probability density function is

$$E[f(t)] = \int_0^\infty tf(t)dt = \frac{1}{\lambda}, \tag{2.8}$$

which is denoted as the mean time to failure (MTTF) of such a component under consideration.

For systems or repairable items, the concept that describes their behavior in terms of being in a failed state or not is named availability. The term availability can be used in different senses depending on the type of system or item under consideration [4]:

1) Availability, A, is the probability that an entity works on demand. This definition is adequate for standby systems.
2) Availability, $A(t)$, is the probability that an entity is working at a specific time t. This definition is adequate for continuously operating systems.
3) Availability, A, is the expected portion of the time that an entity performs its required function. This definition is adequate for repairable systems.

The last definition is the one among the three that represents best the differences between the definitions of reliability and availability. One of these differences was already pointed out and relates to the notion that reliability is a concept that does not apply to systems that may go out of service due to either unexpected or expected causes, and that are brought back to service after some time passed. Another of the differences affecting when the concept of availability needs to be applied originates in the fact that many systems can maintain operation within required parameters even when some of their components are out of service or, after a failure, when not all components that failed have been repaired. As was done for the definition of reliability, it is possible to define a complement to 1 of the availability, which is called unavailability U_a.

Availability calculation can be performed by modeling the failure and repair cycles of a system using a Markov model, shown in Fig. 2.87. In this figure the condition of the system is described by two states indicated by $x(t)$: "working" when it is operating and, thus, meeting its specified operating requirements, or "failed" when the system is

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

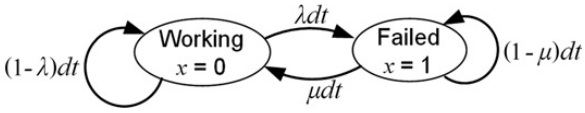

Figure 2.87 Markov process representation of a system with working and failed states.

not achieving one or more of such operating goals. The probability for a repairable item to transition from the working state to the failed state is given by λdt, whereas the probability associated to the converse transition is μdt, where μ is the repair rate. Evidently, the probability of remaining in the working state is given by $(1 - \lambda)dt$ and the probability of remaining in the failed state is $(1 - \mu)dt$. Consider now the third definition of availability given earlier. Then, the instantaneous unavailability of the discussed entity can be associated with the behavior of the item with respect to the failed state $x = 1$. That is, if the probability of finding the entity at the failed state at $t = t + dt$ is identified by $\Pr_f(t + dt)$, then this probability equals the probability that the item was working at time t and experiences a failure during the interval dt or that the item was already in the failed state at time t and it is not repaired during the immediately following interval dt. In mathematical terms,

$$\Pr_f(t + dt) = \Pr_w(t)\lambda dt + \Pr_f(t)(1 - \mu)dt. \tag{2.9}$$

Since $\Pr_f(t) = 1 - \Pr_w(t)$, then it can be found that when $\Pr_f(t = 0) = 0$, then

$$\Pr_f(t) = \Pr[x(t) = 1] = \frac{\lambda}{\lambda + \mu}\left(1 - e^{-(\lambda+\mu)t}\right), \tag{2.10}$$

which implies that

$$\Pr_w(t) = \Pr[x(t) = 0] = 1 - \Pr[x(t) = 1] = \frac{1}{\lambda + \mu}\left(\mu - \lambda e^{-(\lambda+\mu)t}\right). \tag{2.11}$$

Hence, if the system under study had been placed into operation for the first time a long time in the past and since then it has undergone many failure and repair cycles, the steady-state availability and unavailability values are found to equal

$$A = \lim_{t\to\infty} \Pr[x(t) = 0] = \frac{\mu}{\lambda + \mu} \tag{2.12}$$

and

$$U_a = \lim_{t\to\infty} \Pr[x(t) = 1] = \frac{\lambda}{\lambda + \mu}. \tag{2.13}$$

It is now possible to define a mean up time (MUT) as the inverse of the failure rate λ and a mean down time (MDT) as the inverse of the repair rate μ. The MDT includes the processes of detecting the failure, repairing the failure, and putting the item back into operation. The mean time between failures (MTBF) is defined as the sum of the MUT

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

and MDT. With these definitions, the availability and unavailability of an entity can be calculated based on

$$A = \frac{MUT}{MTBF} \tag{2.14}$$

and

$$U_a = \frac{MDT}{MTBF}, \tag{2.15}$$

respectively.

The condition that the system undergoes many failure and repair cycle processes assumed in (2.12) and (2.13) introduces a fundamental distinction between availability and resilience calculations. Because resilience relates to uncommon events, it is not possible to assume that a system will undergo so many multiple service losses and restoration processes during such a disruptive event. Hence, this difference needs to be taken into consideration when defining and calculating resilience, as further explained in Chapter 3.

Markov processes can be used to study systems in which each component is associated to a failed and working state. However, this approach for analyzing such a system becomes very complex because the number of states of the system becomes two to the power of the number of components. One of the alternative methods to represent the availability behavior of a system is through *availability success diagrams*. An availability success diagram is a graphic representation of the availability relationships among components in a system. Such a diagram has the following parts:

a) A starting node
b) An ending node
c) A set of intermediate nodes
d) A set of edges

In the availability success diagram, the edges represent the system components and the nodes represent the system architecture from an availability standpoint. This architecture may be different from a physical or an electrical topology. For example, if the system is an electrical circuit in which there are two components that are electrically connected in parallel but that are critical for the circuit operation – that is, if one of those components fails, the system is in a failed state – then in an availability success diagram they are represented in a series connection. The expected system operating condition is represented by paths through the network. The system is in a working condition when all the components along at least one path from the starting node to the end node are operating normally. If there are enough failed components that it is not possible to find at least one path from the starting node to the end node with all the components operating normally, then the system is in a failed state.

Another method to represent and calculate system availability is the minimal cut sets (mcs) method. An mcs is a group of failed components such that, when all of those components are in a failed state, the system is also in a failed state – characterized in

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

a local area power or energy system (LAPES) by the impossibility of completely feeding the load – but if any single one of those components is repaired, then the system is back again in an operational state. Once the mcs of a system are identified, the unavailability of a system can be calculated from

$$U_a = \Pr\left\{\bigcup_{j=1}^{M_C} K_j\right\}, \tag{2.16}$$

where K_j represents the M_c mcs in the system. Calculating system unavailability using the exact expression in (2.16) is a very tedious process involving identifying the probability of the logical union of many events. However, if all considered components are highly available, then U_a can be approximated to

$$U_a \cong \sum_{j=1}^{M_C} \Pr\{K_j\}, \tag{2.17}$$

where $\Pr\{K_j\}$ is the probability of observing the mcs j happening. Such a probability can be calculated based on

$$\Pr\{K_j\} = \prod_{i=1}^{c_j} u_{i,j}, \tag{2.18}$$

where c_j is the number of failed components in the mcs j, and $u_{i,j}$ is the individual unavailability of each of the c_j components in mcs K_j. Based on (2.13), $u_{i,j}$ is the ratio of the failure rate $\lambda_{i,j}$ of component i in mcs j to the sum of this same component failure rate $\lambda_{i,j}$ and the repair rate $\mu_{i,j}$.

In order to complete the general discussion about availability calculation in systems with multiple components, let's consider some basic systems with commonly found relationships among components. For large systems comprising components arranged in multiples of these structures, it is usually possible to calculate availability characteristics of each of the structures separately and then combine all the structures in order to calculate the total system availability. The three basic commonly used cases are:

1) Series systems

If the system has two components, the availability success diagram is that in Fig. 2.88 (a). For n components, the availability is given by

$$A_{SYS} = \prod_{i=1}^{N_C} a_i, \tag{2.19}$$

where a_i is the availability of each of the N_C components in the system, with the failure rate of the system given by

$$\Lambda_{SYS} = \sum_{i=1}^{N_C} \lambda_i, \tag{2.20}$$

whereas the system repair rate is given by

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

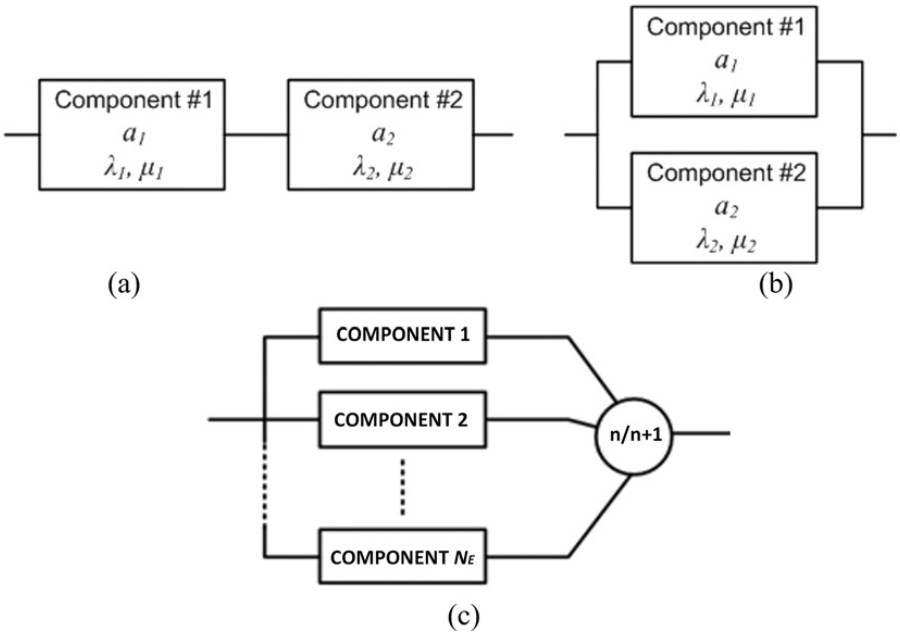

Figure 2.88 Availability success diagrams for (a) two series components, (b) two parallel components, and (c) $n + 1$ redundant components.

$$\mathrm{M}_{SYS} = \frac{\left(\sum_{i=1}^{N_C} \lambda_i\right)\left(\prod_{i=1}^{N_C} \mu_i\right)}{\left(\prod_{i=1}^{N_C} (\lambda_i + \mu_i)\right) - \left(\prod_{i=1}^{N_C} \mu_i\right)}. \tag{2.21}$$

2) Parallel systems

Since in parallel systems there is only one mcs, the expression in (2.17) is now exact, so

$$U_{a,SYS} = \prod_{i=1}^{N_C} u_i, \tag{2.22}$$

where u_i is the unavailability of each of the components. For a two-component system, the availability success diagram is that in Fig. 2.88 (b). As a dual case with respect to the series configuration, the system repair rate is

$$\mathrm{M}_{SYS} = \sum_{i=1}^{N_C} \mu_i, \tag{2.23}$$

and the system failure rate is

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

$$\Lambda_{SYS} = \frac{\left(\sum_{i=1}^{N_C} \mu_i\right)\left(\prod_{i=1}^{N_C} \lambda_i\right)}{\left(\prod_{i=1}^{N_C} (\lambda_i + \mu_i)\right) - \left(\prod_{i=1}^{N_C} \lambda_i\right)}. \tag{2.24}$$

3) $n + 1$ redundant systems

Consider that a system has a number of equal components that all serve the same function. Redundancy is a fault tolerance technique in which the system is equipped with more than the minimum number of these equal components in order to perform their required function adequately and keep the system operating. The most common case of redundancy is the $n + 1$ one in which the minimum number of components necessary to keep the system operating is n and one more component is added as redundancy. Its availability success diagram is represented in Fig. 2.88 (c). Based on the second definition of availability in Section 2.1, system availability is the probability of observing the system to be working. In n +1 redundant systems, such an event – having the system working – is observed when all n +1 redundant components are operating normally or when n of the n +1 components are operating normally. Since there are ${}^{n+1}C_n$ ways in which n operating components can be selected of a group of $n + 1$ components, the availability can be mathematically calculated as

$$A_{SYS} = {}^{n+1}C_n a^n u + {}^{n+1}C_{n+1} a^{n+1}, \tag{2.25}$$

where a and u are the availability and unavailability, respectively, of the $n + 1$ equal components in the $n + 1$ redundant arrangement and where

$${}^{k}C_n = \binom{n}{k} = \frac{k!}{(n-k)!n!}. \tag{2.26}$$

Hence,

$$A_{SYS} = (n+1)a^n u + a^{n+1}. \tag{2.27}$$

When (2.27) is studied, it is possible to observe that as the minimum number of components n is increased, the system availability decreases; and for values of n large enough, the system availability A_{SYS} is less than the individual component availability a. Hence redundancy is a fault tolerance technique that needs to be used with care, as increasing the number of components may compromise system availability instead of improving it. Finally, the failure and repair rates in an $n + 1$ redundant arrangement are given by [4]

$$\Lambda_{SYS} = \frac{n\lambda^2(n+1)}{(n+1)\lambda + \mu} \tag{2.28}$$

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

and

$$M_{SYS} = \frac{2({}^{n+1}C_{n+1})\lambda^2\mu^n}{\sum_{i=0}^{n-1}\left({}^{i}C_{n+1})\mu^i\lambda^{n+1-i}\right)}, \tag{2.29}$$

respectively.

As discussed in Chapter 4, service buffers mitigate the negative effect of dependencies on resilience. Energy storage devices, such as batteries, are the realization of buffers for the electric power provision service. Hence, it is relevant to understand how energy storage is taken into consideration when calculating availability. As explained in [5], electric power service unavailability, U_S, with an energy storage backup able to keep the load powered for a time equal to T_S is given by

$$U_S = U_G e^{-\mu_G T_S}, \tag{2.30}$$

where U_G is the unavailability of the electric power provision service without energy storage backup and μ_G is the repair rate for the electric power provision service.

Markov process theory also allows for calculating the ac power supply availability at the rectifiers' input in Fig. 2.62, assuming an ideal fuel supply for the genset; that is, fuel supply availability for the genset equals 1. Based on this assumption, from [6] and [7], ac power supply availability is given by

$$A_{ac} = \left(1 - \frac{(\lambda_{GS} + \rho_{GS}\mu_{MP})\lambda_{MP}}{\mu_{MP}(\mu_{MP} + \mu_{GS})}\right)A_{TS}, \tag{2.31}$$

where A_{TS} is the transfer switch availability, λ_{GS} is the failure rate of the series combination of the generator set, μ_{GS} is the genset and fuel repair rate, ρ_{GS} is the genset failure-to-start probability, λ_{MP} is the mains power failure rate, and μ_{MP} is the mains power repair rate. Later chapters of this book explain approaches for calculating resilience when using diesel-fueled electric generators without assuming that fuel supply availability is 1.

During operations under normal conditions, electric power distribution utilities in the United States use various metrics in IEEE Standard 1366 [8] in order to evaluate their service reliability – this concept is now understood within the broader scope of an engineering field. These metrics in [8] are calculated excluding outages occurring in what is defined as a "major event day." That is, outages caused by natural disasters and other extreme events are not considered as part of the calculation of the metrics, which explicitly indicates that such indices are not applicable for calculating resilience. Thus, although there are differences between resilience and reliability or availability metrics, it is still relevant to indicate some of the metrics indicated in this standard. Consider first metrics concerning sustained interruptions – those lasting more than five minutes:

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

– System average interruption frequency index (SAIFI)

$$\text{SAIFI} = \frac{\sum \text{Total Number of Customer Interrupted}}{\text{Total Number of Customers Served}} \tag{2.32}$$

That is, the numerator equals the sum of the "number of interrupted customers for each sustained interruption event during the reporting period" [8].

– System average interruption duration index (SAIDI)

$$\text{SAIDI} = \frac{\sum \text{Customer Interruption Durations}}{\text{Total Number of Customers Served}} = \frac{\sum r_i N_i}{N_T}, \tag{2.33}$$

where r_i is the "restoration time for each interruption event," N_i is the "number of interrupted customers for each sustained interruption event during the reporting period" and N_T is the "total number of customers served for the areas" under consideration [8].

– Customer average interruption duration index (CAIDI)

$$\text{CAIDI} = \frac{\sum r_i N_i}{\sum N_i} = \frac{\text{SAIDI}}{\text{SAIFI}}, \tag{2.34}$$

– Average service availability index (ASAI)

$$\text{ASAI} = \frac{N_T T_{H/Y} - \sum r_i N_i}{N_T T_{H/Y}}, \tag{2.35}$$

where $T_{H/Y}$ is the number of hours in a year (8,760 in a nonleap year and 8,784 in a leap year). That is, the ASAI "represents the fraction of time that a customer has received power during the defined reporting period" [8].

– Customers experiencing multiple interruptions (CEMI_n)

$$\text{CEMI}_n = \frac{CN_{k>n}}{N_T}, \tag{2.36}$$

where $CN_{k>n}$ is the total number of customers experiencing more than n sustained interruptions.

Other metrics in [8] are based on other evaluation parameters, such as the load's power consumption. These are:

– Average system interruption frequency index (ASIFI)

$$\text{ASIFI} = \frac{\sum L_i}{L_T}, \tag{2.37}$$

where L_i is the connected load apparent power (kVA) "interrupted for each interruption event" [8] and L_T is the total load apparent power (kVA) served.

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

– Average system interruption duration index (ASIDI)

$$\text{ASIDI} = \frac{\sum r_i L_i}{L_T}, \tag{2.38}$$

Finally, some metrics in [8] are specified for momentary interruptions. They are as follows:

– Momentary average interruption frequency index (MAIFI)

$$\text{MAIFI} = \frac{\sum I_{Mi} N_{mi}}{N_T}, \tag{2.39}$$

where I_{Mi} is the "number of momentary interruptions" and N_{mi} is the "number of interrupted customers for each momentary interruption event during the reporting period" [8].

– Momentary average interruption event frequency index (MAIFI_E)

$$\text{MAIFI}_E = \frac{\sum I_{ME} N_{mi}}{N_T}, \tag{2.40}$$

where I_{ME} is the "number of momentary interruptions events" [8].

– Customers experiencing multiple sustained interruption and momentary interruption events (CEMSMI_n)

$$\text{CEMSMI}_n = \frac{CNT_{k>n}}{N_T}, \tag{2.41}$$

where $CNT_{k>n}$ is the "total number of customers who have experienced more than n sustained interruptions and momentary interruption events during the reporting period" [8].

2.5 Disruptive Events

There are many disruptive events that can affect the operation of power grids, information and communication networks, and other critical infrastructures, so this section discusses those that are considered the most relevant ones. Disruptive events can be broadly classified between natural disasters and man-made events.

2.5.1 Natural Disasters

As the term indicates, natural disasters are events that occur without direct human intervention. Because of their natural origin, natural disasters cannot be generally prevented, although they can be anticipated with a given probability distribution, and

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

some can be forecast with a reasonable degree of accuracy some time – usually days – in advance. The following are relevant natural disasters:

– *Tropical cyclones*: These extremely intense storms, also known as hurricanes in North America or typhoons in eastern Asia, typically affect tens of thousands of square kilometers, and are not limited to coastal areas where the effects are more intense. Cyclones' damaging actions include very intense winds, inland flooding, and coastal storm surge, which is an influx of seawater carried inland by the storm by a combination of its strong winds and low pressure. The most intense winds are usually observed in a relatively small area near the "eye" or center of circulation of the storm. Typhoons' and hurricanes' intensity is commonly classified based on their maximum sustained winds. For example, the Saffir–Simpson scale used by the US National Hurricane Center, calls tropical storms those that have tropical origin and have one-minute-average maximum sustained winds at 10 m above the sea surface between 39 and 73 miles per hour (mph). Hurricanes are storms with one-minute-average maximum sustained winds at 10 m above the sea surface of more than 73 mph. Category 1 hurricanes are those with such winds between 74 and 95 mph, category 2 hurricanes have winds between 96 and 110 mph, category 3 hurricanes have winds between 111 and 130 mph, category 4 hurricanes have winds between 131 and 155 mph, and category 5 hurricanes have winds greater than 155 mph. The main goal of this scale and other similar ones, such as the one used by the Japan Meteorological Agency, is to be simple so that it conveys an idea of the storm's intensity to the general public. However, there are other factors, such as storm surge height, that also affect critical infrastructure's resilience and, thus, need to be taken into account when evaluating these storms' impact on infrastructures. Thus, in 2007 [9] suggested an alternative measure of hurricane intensity called the Integrated Kinetic Energy (IKE). Figure 2.89 shows an example of extreme storm surge damage. Such extreme destruction is relatively unusual and, as with many natural disasters' damaging actions, such intense effects are limited to a relatively much smaller area over the entire region affected by the extreme event. That is, it is relatively common to observe areas with much less damage or even no significant observable damage some hundred meters away from the more intense damaged

Figure 2.89 Example of intense storm surge damage.

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

Figure 2.90 Storm surge damage to buildings may be different from effects on built infrastructure components.

area. Additionally, storm surge only affects coastal areas and thus its effects are dependent on the geographic characteristics of the coast, such as change of elevation and shape. Tropical cyclones' strongest winds are also observed in coastal areas because these storms weaken over land. It is also important to point out that damage extent and severity to dwellings and other buildings may be different from that observed to infrastructure systems, as is exemplified in Fig. 2.90. There are also other storm factors that do not necessarily cause damage but still affect resilience because of their effects on restoration activities. Examples of these factors include size of the storm and time under at least tropical storm winds [10]–[11]. Its displacement speed is another important factor because slow-moving storms tend to increase the chances of flooding due to persistent precipitation. In addition to floods, it is common that tropical cyclones originate tornadoes and intense hail. Current weather models allow one to predict the general region that is affected by these storms a few days in advance.

– *Earthquakes*: These are manifested by shaking of the earth's surface as a result of waves, called seismic waves, caused by the release of energy occurring most commonly when geologic faults rupture. Because earthquakes usually originate due to geologic fault ruptures, they are typically observed where tectonic plates collide. There are, however, other origins of earthquakes, such as volcanic eruptions and nuclear explosions. A primary damaging action of earthquakes is, thus, the mechanical stresses caused on infrastructure systems' components due to the shaking. Examples of this type of damage include broken bushings and insulators of high-voltage electrical equipment, transformers losing their anchoring, damage to batteries and gensets, and antennae misalignment. High-voltage transformers are

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

Figure 2.91 Damage caused on poles by debris carried by tsunami waves.

usually automatically disconnected even when they experienced no damage when shaking caused the oil level of these transformers to trip their Buchholz relays. However, other damaging actions can affect infrastructure systems' resilience. Ground failure, land- or rock-slides, and soil liquefaction – when soils lose their solid characteristic and become a liquid due to shaking – may cause damage to both buried and aboveground components (e.g., see Fig. 2.39 (a) and additional images in Chapter 5) and they may affect restoration logistics due to damaged or obstructed roads. Additionally, earthquakes with their epicenter occurring on a large body of water, such as a sea or ocean, may cause a tsunami, which is water waves radiating from the epicenter. These waves could be very damaging to both residential and commercial buildings and to infrastructure system components. The natural damaging action of waves is typically compounded by the debris they carry, increasing the damaging potential, for example, over poles and towers of both electric power grids and communication networks. Figures 2.91 and 2.92 exemplify such types of damage. If fires are ignited during the earthquake, a tsunami may carry floating, burning debris, creating an additional damaging action, as exemplified in Fig. 2.93. However, as Fig. 2.94 exemplifies, tsunami waves may not necessarily cause damage when they do not carry debris, as exemplified by wind turbines operating on the Japanese eastern coast during the 2011 earthquake and tsunami that affected that area. In this particular site, shaking was a more important concern because of the high center of gravity and relatively shallow foundation of wind turbines, as exemplified by Fig. 2.95, showing a ring added to a wind turbine column to compensate the tilt resulting from the earthquake shaking. Flooding and soil scour are other damaging actions from tsunamis, as exemplified in Fig. 2.96. After an earthquake, coastal areas may also experience periodic flooding during normal high tides as a result of land subsidence. Earthquakes may also cause fires in urban areas due to ruptured natural gas distribution pipes or due to damage to other heating means using fuels, such as broken connections between propane gas and the house

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

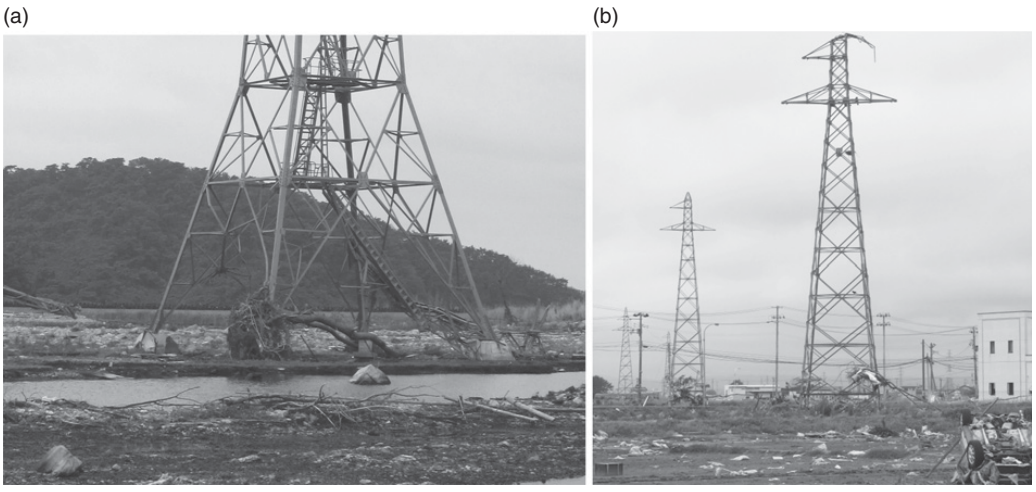

Figure 2.92 Damage to communications tower (a) and braces of electric power transmission towers (b) by uprooted trees and other debris carried by tsunami waves.

Figure 2.93 Fire damage caused by burning debris carried by tsunami waves.

or building gas pipes. Thus, natural gas distribution systems may automatically get disconnected during earthquakes to prevent such fires occurring, and their service is not restored until inspections are completed. This interruption of natural gas service may then affect operation of electric power generators using such fuel. Earthquake magnitudes used to be indicated based on the Richter scale, but nowadays earthquake magnitude is characterized using the moment magnitude – indicated as Mw – which is related to the total energy released by the earthquake and thus is

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

Figure 2.94 Wind turbines undamaged by tsunami waves in Japan after the 2011 earthquake.

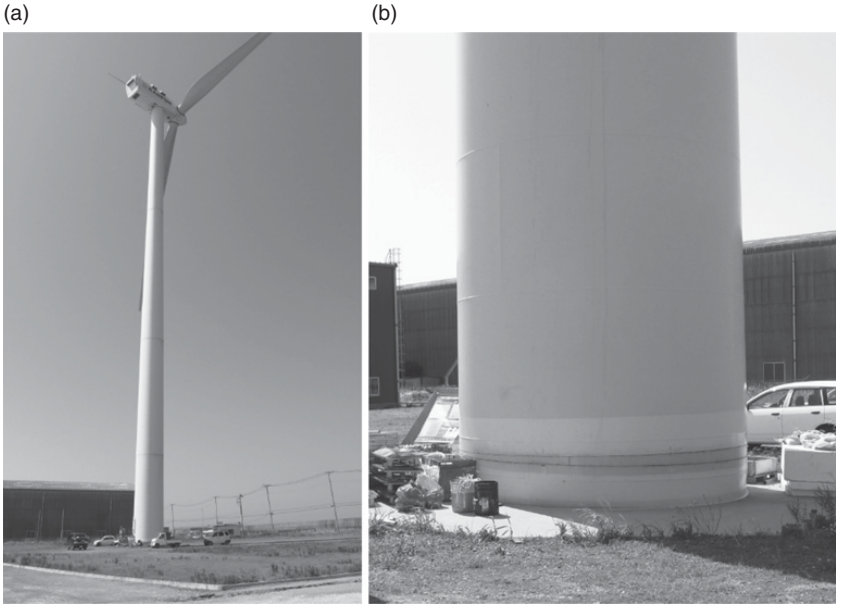

Figure 2.95 A wind turbine that tilted during the 2011 earthquake in Japan. A ring was added at the base to compensate for such tilt (see (b)).

independent of the location of the observer. Currently, the earthquake with the highest observed moment magnitude is the 1960 earthquake that affected the region of Bio Bio in Chile with a moment magnitude equal to 9.5. Intensity of an earthquake measures its shaking at a local level. In the United States the modified

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

Figure 2.96 A destroyed building partially sunk due to ground scour caused by tsunami waves.

Mercalli (MMI) scale is used to characterize earthquake intensity based on perceived observed shaking intensity. Thus, this measurement tends to be subjective. However, there are objective measurements used to measure shaking intensity. A common such measurement is the peak ground acceleration (PGA) that measures the maximum ground acceleration with respect to the acceleration of earth's gravity, g. The US Geological Survey has developed a scale that relates PGA ranges to perceived shaking in the Mercalli scale. Some of the largest recorded PGAs include a value of 2.7 during the 2011 Mw 9.0 Tohoku Earthquake in Japan and of 2.2 during the February 2011 Mw 6.2 Christchurch Earthquake in New Zealand. Although earthquakes can be anticipated with a calculated probability over an indicated period of time of usually many years, it is still not possible to forecast when and where they will happen. The only warning of an earthquake that it is possible to have nowadays is based on detecting the earthquake primary waves that travel faster and are less destructive than the secondary waves. These earthquake warning systems provide from a few seconds up to several tens of seconds of warning of an impending earthquake depending on the distance to the hypocenter.

– *Floods:* These are, arguably, the most common natural disruptive event affecting communities in general. Floods could be caused in various ways. One of these ways is through torrential rains leading rivers and small reservoirs to overflow beyond

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

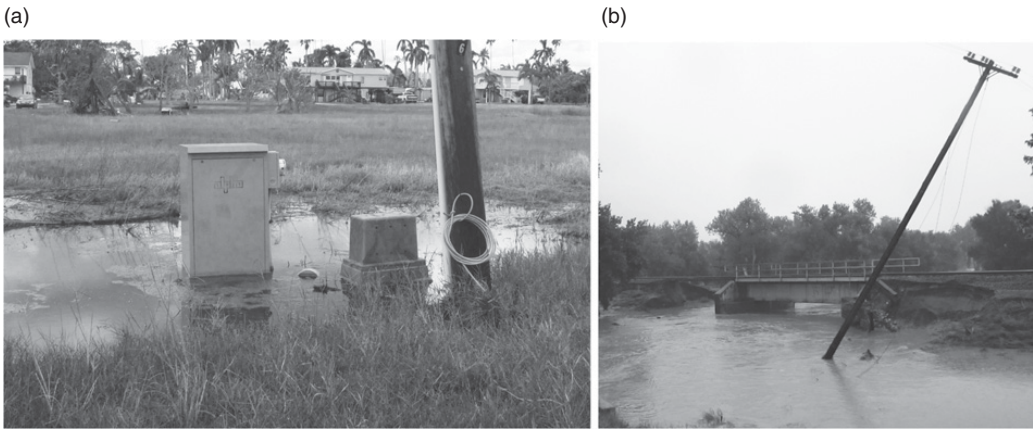

Figure 2.97 Flood damage. (a) Direct damage to a CATV UPS. (b) Tilted pole due to water-saturated soils.

their normal banks within a relatively short time of at most a few hours. Another way is through less intense but sustained rains, which saturate soils with water and also lead rivers to overflow. Another way in which floods may be caused in a region even when such an area is not subject to torrential or sustained rains is for intense rains upstream that eventually make it to regions downstream. Similarly, regions downstream of a river may flood in spring when snow and ice thaw in areas upstream. Generally, even when water damages electrical and electronic equipment, floods tend to be, in the short term, less damaging to power grids and communication networks than other disruptive events. However, floods may create logistical issues from roads or bridges being washed away or made impassable using normal means. Additionally, in mountainous regions floods may cause *landslides* that could cause more damage to infrastructure components than just the damage originating in the effect of water alone. In addition to direct water damage, as exemplified in Fig. 2.97 (a), floods saturate soils with water that could make poles tilt or fall, as shown in Fig. 2.97 (b). Like storms, floods can be anticipated sufficiently accurately a few days before they occur. Moreover, chances of experiencing floods within a given time period have been relatively well characterized in many regions in the world. For example, in the United States, the government produces maps with various zones representing the probability of experiencing a given flood intensity over a 30-year period, which is the normal duration of standard mortgages. The zones in these maps are the ones corresponding to a 25-year flood zone with a 71 percent chance of being flooded, the 50-year flood zone with a 45 percent chance of being flooded, the 100-year flood zone with a 26 percent chance of being flooded, and the 500-year flood zone with a 6 percent chance of being flooded. This information is typically used by infrastructure operators to implement flood mitigation strategies and technologies, such as raising infrastructure components on platforms over a given level indicated by the

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

Figure 2.98 Example of damage caused by a tornado. (a) Destroyed dwellings next to a seemingly undamaged transmission line. (b) An area completely destroyed by a tornado.

flood zone, where it is deemed appropriate based on planning studies, such as those described in Chapter 9.

– *Severe storms:* In addition to floods, which have already been commented on, lightning and hail are potential damaging actions resulting from severe storms. Power grid components, such as overhead transmission lines and aboveground substations, are protected with ground wires to conduct lightning discharges to ground (e.g., see Figs. 2.23 and 2.38). Additionally, surge arresters are used to protect against induced voltages that may still occur even when the lightning strikes on ground conductors or in the vicinity of electric infrastructure both aboveground or buried. Communication network components are also protected with lightning rods and surge arresters, for example, in wireless network towers. In general, except for photovoltaic systems, hail tends to produce less damage to both power grid and communication network components. Grounding systems are a fundamental component of atmospheric discharge protection systems. Such systems normally have a complex design in which a mesh of electrically connected rods buried under the protected facility is connected to a relatively large conducting bar using a low-resistance conductor. All electronic and electrical equipment is then connected to the large conducting bar to ensure almost equal ground potential for all equipment in the facility even when atmospheric discharges occur. Design of the grounding system is influenced by the local soil resistivity. Additionally, grounding systems need to be maintained to ensure an adequate resistance to ground even when soil conditions change. Although the exact location where lightning will strike still cannot be anticipated, forecasts for storms producing significant amounts of lightning or large hail are well developed and have a general good accuracy even one or two days in advance. Additionally, there are well-established climate models able to characterize general patterns of occurrence of this type of severe storms for every month.

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

Figure 2.99 Different damage intensity along a tornado path.

– *Tornadoes*: These are also weather phenomena observed in some intense storms. Tornadoes are rapidly rotating columns of air. Because of their rapid rotation, they generate very strong winds that may cause damage directly due to their intense pressure or indirectly from impacting flying debris, which could include objects as heavy as a car. Thus, tornadoes could produce significant damage to critical infrastructures from these actions, but their damage path is usually characterized by a well-defined boundary, as an abrupt difference in damage intensity is usually observed between the tornado damage path and the neighboring undamaged areas. Because of this abrupt difference between damaged and undamaged areas, exemplified in Figs. 2.98 and 2.99, it is difficult to plan and design mitigation measures due to the highly uncertain probability of being in the damage path. However, weather forecasts are able to provide a few days in advance relatively accurate predictions of observing a given number of tornadoes of an indicated intensity range within 25 miles of the forecast location. Climate models also provide generally accurate estimates of monthly or seasonal trends for intense storms that could generate tornadoes. Tornado intensity is characterized based on the Enhanced Fujita scale, which is divided into six levels, from the least intense one identified by EF-0 with 3-second wind gusts between 65 and 85 miles per hour to the most intense level identified by EF-5 with wind gusts over 200 mph. Typical tornado damage paths have a width of 50 meters and a length of one or two miles. However, the damage path of very intense tornadoes could be tens of miles long and have a width of one mile.

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

– *Ice or snow storms:* Winter storms may have a significant impact, particularly on power grids. A common effect of these storms when they are accompanied by heavy ice or snow accumulations is broken wires. Lines could often be damaged during these storms by fallen trees or branches, also due to excessive snow accumulation. Damage is usually made worse by difficulties in restoration activities due to icy roads, snow accumulation, or working under very cold conditions. These storms may also have both direct and indirect effects on electric power generation by, for example, having ice obstructing power stations' water intakes. Direct effects are observed particularly in renewable energy sources. For example, frozen rivers or floating ice may affect hydroelectric power plants, snow may obstruct sunlight from reaching photovoltaic cells, and ice may reduce wind power generation if wind turbines lack heating elements on their blades. Indirect effects include higher fuel costs and, in some cases, fuel shortages in natural gas power plants due to increased use of the same fuels by other industries and residences for their heating needs. *Cold waves* – various days of extremely cold temperatures – are another form of winter event that could accompany ice or snow storms. Cold waves could also indirectly impact power generation due to their effects on increased fuel demand and electric power consumption. These winter storms are commonly forecast various days in advance with relatively good accuracy.

  Droughts: Droughts are abnormally long periods of significantly lower than normal rainfall causing shortage of water. Typically, droughts affect very large regions. Because of their sometimes-long duration, droughts could be considered climate events. Droughts increase the likelihood of wildfires. Droughts are usually accompanied by higher-than-usual temperatures causing higher demand for electric power for air conditioning systems. Additionally, various *heat waves*, when temperatures for a few days are significantly higher than usual, are often associated with droughts. During these heat waves, power consumption increases even more due to the increased use of air conditioning systems. The most direct impact of droughts is on hydroelectric power plants as reservoir levels decrease, as exemplified in Fig. 2.100. Cooling of electric power generation stations and of information and communication network facilities can also be affected during droughts if use of water is rationed. Additionally, loss of humidity in soils may cause grounding issues due to soil changing electrical conductivity. Long and intense droughts may eventually cause issues with foundations and buried infrastructure components when soils dry up and crack. A particularly challenging aspect of droughts is that, although their onset can be anticipated, it is very difficult to predict their duration and intensity.
– *Wildfires:* These are uncontrolled and unplanned fires affecting areas with vegetation. Thus, they commonly originate in rural areas and can be caused naturally (e.g., by lightning) or by humans either intentionally or accidentally. Wildfires can sometimes affect large areas, but because they usually start in rural areas their direct impact on electric or communication infrastructures through

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

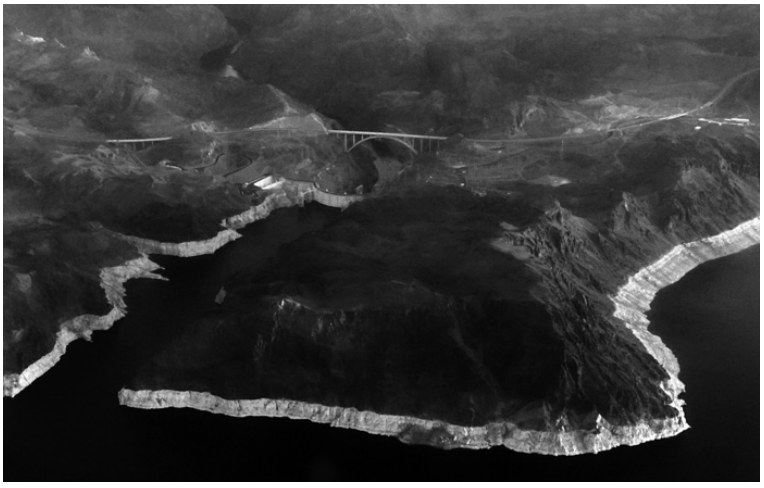

Figure 2.100 Significant drop in the Hoover Dam reservoir water level during the drought that affected the Colorado River in 2021.

damage is limited even if wildfires move to urban areas because areas with vegetation where wildfires could be more intense and spread quicker are also usually less populated areas where there is not a dense deployment of infrastructure components. Still, wildfires could cause some noticeable electric power outages if the fire affects transmission lines or if a sufficiently large number of these lines need to be taken out to facilitate fire extinguishing activities. However, wildfires can have a significant indirect effect on electric utilities due to the economic liabilities that these companies could face if wildfires are proved to have started from some issue in their grid, such as fires initiated from sparks when an energized conductor falls to ground, causing a short circuit. Such economic impact could lead to electric utilities bankruptcy, as exemplified by the case of Pacific Gas and Electric in the state of California in the United States. Because conditions when wildfires are more likely to occur, such as dry and windy weather, can be forecast, during recent years electric utilities in states such as California or Oregon have been implementing preventive electric power outages when those weather conditions are forecast. In some cases, such as in October 2019, these preventive power outages affected almost one million users for a few days. Evidently, these preventive power outages also have an economic impact due to lost revenue, but this impact is less costly than the potential costs related to wildfire liabilities.

– *Volcanic eruptions:* Volcanic eruptions are the release of lava and/or gases from volcanoes. Even when eruptions do not release large amounts of lava, they could expel tephra – a term describing ash, rocks, and other material ejected from the volcano. Volcanic eruptions are commonly associated with tremors. Some volcanic eruptions could be explosive. Since it is uncommon that volcanoes are surrounded

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

by heavily populated areas with densely built infrastructure systems, it is relatively unusual to observe extensive damage to electric power grids or communication networks from explosive eruptions or from lava or pyroclastic flows because of the few infrastructure components in these areas. Still, it is possible to find examples of lava or pyroclastic flows causing extensive damage to infrastructure systems during past volcanic eruptions, such as the Soufrière Hills volcano eruption in 1995. However, tephra fallouts could cause important issues with electric power grids and communication networks over a relatively large area. Various types of damaging action by tephra on electric power grids are described in [12]. These actions include accelerated wear of hydroelectric power plant turbines due to tephra's abrasive nature and insulator flashovers because wet volcanic ash conducts electricity. Even if flashovers do not occur, tephra-induced leakage currents may degrade insulating characteristics of insulators. Other effects of tephra deposits include fuel contamination and air filter obstruction of combustion-driven generators, grounding issues due to reduced soil resistivity, and obstructed cooling vents of electrical machines. Additionally, tephra causes long-term accelerated corrosion if it is not completely removed from unprotected metallic surfaces, which, evidently, it is practically impossible to do, as winds keep on dispersing ash from surrounding areas and depositing it after cleanup is completed. Restoration activities could also be affected during volcanic eruption due to damage or obstructions to roads. Volcanic eruptions can be anticipated months in advance, and seismic activity and other monitoring tools are commonly used to predict when an eruption could happen within the following few days.

– *Geomagnetic storms* (GMS): Despite their name, these disruptive events are not Earth-atmospheric weather events but, instead, they are space weather–induced events. Geomagnetic storm is the name given to significant changes in Earth's magnetic field caused primarily by interactions between Earth's magnetosphere and the Sun. The Sun is continually emitting a stream of high-energy charged particles – mostly electrons and protons – called solar wind, which is an electric current that interacts with the Earth's magnetic field, causing several phenomena including auroras and geomagnetic-induced currents. A geomagnetic storm is generated when the solar wind intensifies, usually due to processes originating in corona holes or due to corona mass ejections (CMEs) [13]. Corona holes are areas in the Sun's corona with lower density of plasma and where charged particles can escape the Sun easier due to the particular configuration of the magnetic field. Corona holes are the most common source of solar wind–inducing GMSs, but these storms have a more gradual onset and are milder than those created by CMEs [14]. Coronal mass ejections are a sudden release of Sun plasma primarily formed by electrons and protons from the Sun's upper layers in areas near sunspots, which are dark and colder regions on the Sun's surface generated by particular local configurations of the Sun's magnetic field. This sudden release of electrically charged matter creates a significantly abrupt increase in the solar wind, which is more intense, as it contains more charge particles moving at higher velocities. Although very intense CMEs can reach Earth in 15 to 18 hours, most common CMEs take two or three days to reach Earth. When the solar wind reaches

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

Earth's proximity, its intrinsic magnetic field directs the charged particles toward the polar regions, following a spiral trajectory along Earth's magnetic field intensity lines. When the solar wind particles reach the Earth's atmosphere, they produce electrojets – horizontal currents in Earth's atmosphere circulating around magnetic poles and with intensities that could reach as high as a million amperes. In normal conditions, the Earth's magnetic field remains steady, not being severely affected by the solar wind, and the only notable effect is the generation of auroras visible at night in high-latitude regions. However, when the solar wind is strong enough, for example, as a product of a CME, it generates electrojets with enough intensity to disrupt Earth's intrinsic magnetic field by making it vary. This varying magnetic field induces voltages along Earth's surface of about 1.2 to 6 V/km that in turn generate geomagnetically induced currents (GICs), also on Earth's surface [15]. These GICs are characterized by their low-frequency components, ranging from about half a hertz to as low as a thousandth of a hertz [16]. The varying magnetic field is used to characterize GMSs' intensity through the Kp index. This index is obtained based on the weighted average K-index measured at several locations, where the K index is "the maximum fluctuations of horizontal components observed on a magnetometer relative to a quiet day, during a three-hour interval" [17]. Although CMEs can occur at any time when the conditions on the Sun are suitable for their formation, their occurrence follows a varying profile, with the maximum number of occurrences happening approximately every 11 years. Interactions of GICs with built infrastructures and, in particular, electric power systems and communication networks have been documented for more than 100 years [16], [18]–[25]. The most notable effects for power systems are observed at the power grid's transmission level because long lines running primarily in an east–west direction and over high-resistivity terrain make them susceptible to facilitate geomagnetically induced voltages to appear [26], which would, in turn, generate GICs when a loop is created in some particular circuit configurations, specifically when the line is terminated at both ends in wye-connected transformer windings with their center grounded [16]. Since GICs are quasi dc currents [16], even relatively small currents can drive the transformer cores to half-cycle saturation, causing in turn high content of odd- and even-baseband harmonics, higher eddy-current losses, voltage unbalances, and higher reactive power consumption that could lead to undesirable voltage drops [16]. During intense GMSs, additional transformer losses could be high enough to make transformers fail due to excessive heating [16]. However, even when the GICs are not intense enough to lead to such catastrophic failure, repeated exposures to moderate GMSs may degrade transformer winding insulations, which will shorter the transformer's life [16]. Although it is commonly believed that the effects of GMSs are limited to high-latitude regions – where latitude refers to magnetic latitude, which may not coincide with the geographic latitude – minor to moderate effects have been documented in the United States as far south as central California and north Texas [23] [26]–[28]. Moreover, it is believed that during extremely powerful CMEs severe effects could be observed in the entire contiguous United States and, of course, Alaska. Such severe storms do not need to duplicate the one called the Carrington Event of 1859, when a powerful CME triggered auroras that

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

could be seen as far south as 23 degrees in latitude [29]. Severe GMSs half the intensity of the Carrington event but with sufficient intensity to create auroras and GICs severe enough to cause damage in power systems in low- latitude regions are expected to occur every 50 years [30]. Past records of important GMSs include those in 1872, 1921, 1938, 1940, 1958, 1960, 1972, 1989, and 2003. Geomagnetic storms can directly affect power grids at the transmission level in other ways. Increased use of electronically controlled protective relays makes grids more prone to failure under the presence of higher harmonic content in the power signal [16]. Reactive power compensators and shunt capacitor banks can trip due to currents along their neutral grounded connection [16]. A few reports indicate that some surge arresters failed in the past due to neutral overvoltages [13]. Power-line carrier communications are also negatively affected by GMSs because GIC and higher system harmonic content cause a decrease in signal-to-noise ratio [25]. Geomagnetic storms can also affect communications. The most immediate effect is radio blackouts on Earth's side facing the Sun shortly after an intense CME is ejected. These radio blackouts are caused by ionization of the lower layers of the ionosphere of Earth's side facing the Sun from increased levels of X-ray and ultraviolet electromagnetic radiation produced by solar flares typically accompanying CMEs – solar flares are large eruptions of electromagnetic radiation from the Sun. CMEs may have a considerable impact on satellite operations due to damaged electronics from increased radiation or from disruptions of Earth's atmosphere and magnetosphere. In particular, CMEs increase the temperature of Earth's outer atmosphere, causing it to expand. As the atmosphere expands, drag on low-Earth-orbiting satellites increases, thus reducing their lifetime. Geomagnetic storms can also induce currents on long communications wire lines. However, these types of lines are almost not used anymore in practically all modern communication networks. Although it is still not possible to forecast the exact moment when CMEs will happen, data from sun-monitoring satellites allows one to specify chances of such CMEs happening within the following two or three days. Moreover, equipment in these satellites can observe when and how CMEs happen and anticipate whether such CMEs may be Earth directed and, in such a case, their arrival time. Milder geomagnetic storms caused by coronal holes tend to be simpler to forecast, based on satellite data monitoring the Sun.

– *Pandemic:* A pandemic is said to exist whenever an infectious disease spreads over a large area (which may be the entire world), affecting a very large number of people. A disease with a relatively steady number of cases is not considered a pandemic, even if it also extends over a very large area – such a type of disease is called endemic. Pandemics are considered natural disruptive events even if in some cases their origin could be man-made. For example, a biological attack becomes a pandemic when it begins to spread over a large region, infecting a large number of people and thus the disease-spreading mechanism, which is a main characteristic of a pandemic, is a natural process, meaning that pandemics are better defined as natural disruptive events. Evidently, pandemics do not cause damage to critical infrastructures. However, they can affect their operations both directly when infrastructure operators get infected and require medical care and possibly

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

quarantining, and indirectly through the negative economic impact associated with intense and extensive infectious events. Pandemics may not be able to be forecast, but it is possible to plan mitigation measures in the organizational processes used to manage and administrate utilities and other business operating critical infrastructures.

2.5.2 Human-Driven Disruptive Events

Human-driven disruptive events are those in which humans play, through their decisions and actions, the main role in not only originating the event but also in some instances keeping such an event continuing. Human-driven disruptive events could take different forms. One type of human-driven disruptive evens could be attacks on infrastructure components or on other targets but also affecting infrastructure system components. The most common example of attacks is *explosive attacks* or sabotages. These types of attacks are common in both conventional or asymmetric – namely, guerrilla or insurgence warfare – armed conflicts. Power grids are common targets during conflicts, and although a main concern is attacks on substations, as happened with the sniper attack on the substation shown in Fig. 2.35, or on power plants because of their importance due to power grids' centralized architectures, the most common and simplest target of such attacks is transmission lines because of the practical impossibility in protecting all transmission lines in their entire extension. *Extreme types of attacks* during armed conflicts are those using *weapons of mass destruction* (WMDs): nuclear, bacteriological, or chemical devices. While these last two extreme types of attacks do not cause damage to infrastructure components and only affect people, nuclear attacks cause destruction and damage from blast pressure [31]–[32], intense heat and fire propagation, and may also affect people long after the bomb explodes due to radioactive elements' fallout and contamination. Evidently, power grids and other critical infrastructures may likely be severely affected during these types of attacks [33]–[36]. Moreover, large ICN facilities, particularly telecommunications central offices, could still experience severe damage even if they survive the explosion, as exemplified in Fig. 2.101. Additionally, nuclear explosions could create electromagnetic pulses (EMP), which are electromagnetic disturbances with some similarities to geomagnetic storms that could cause service disruptions to power grids and ICNs, as demonstrated during the Starfish Prime nuclear test in 1962 [37]. *Cyber-attacks* are another type of human-driven disruptive event that have recently been attracting increased interest. The target in cyber-attacks is the computing and information assets (including databases, sensing subsystems, and control algorithms) used to monitor and control infrastructure systems. This type of attack requires, however, a larger team with a significant amount of education and training to gain the necessary extensive computer systems and software knowledge than a much-reduced team of attackers with little training or education other than simple knowledge about using explosives or conducting basic sabotage activities. However, cyber-attacks on critical infrastructure systems may be an attractive approach for nations to subvert

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

Figure 2.101 NTT West telephone central office in the city of Hiroshima as seen today and two months after the nuclear explosion in 1945 on the plaque at the bottom right of this image. Although this central office was one of the very few surviving buildings within a one-mile radius from the explosion, it still sustained damage to its equipment.

foreign countries because these types of attacks may be conducted from abroad and may also be seen as less harmful than conventional attacks, which could be easily considered an act of war, whereas a cyber-attack is significantly more difficult to be considered in such a way. All types of attacks can usually be anticipated, based on the political and security conditions existing where the power grid is operating.

Another significant human-driven disruptive event is an *economic crisis*. As discussed in more detail in future chapters, economic crises are significant disruptive events for critical infrastructures. Although economic crises do not cause immediate damage to infrastructure components, they reduce resilience by impacting preparedness activities, such as training for reducing restoration times, in a different future disruptive event. An example of the impact of economic crisis is found in Puerto Rico, where the electric power grid was already weak as a result of an economic disruption that even caused the local electric utility to file for bankruptcy protection some time before the hurricane affected the island [38]. Economic disruptions could be caused by various causes, including direct or indirect effects of government policies and regulations and business mismanagement. Recovery from economic disruption depends on many factors, including potential financial assistance from the government or general economic conditions. Thus, although in many instances economic disruptions could be anticipated a few months before they potentially materialize, it is extremely difficult to forecast when such economic disruptions end. Moreover, even when economic

https://doi.org/10.1017/9781108648899.003 Published online by Cambridge University Press

disruptions could be anticipated months before their effects could be felt, it may not be possible to implement mitigation or corrective actions in such a timeframe.

One other possible disruptive event related to human actions is *technology disasters*. These events can have different origins, including policies, standards, design or planning issues, and very unlikely technological accidents with a high impact. Notice that only accidents that are very unlikely and with a high impact can be considered in resilience studies. One example of a technological disaster originating in unlikely accidents is a very serious nuclear accident, such as those in the top two magnitude levels according to the seven-level International Atomic Energy Agency's (IAEA) International Nuclear and Radiological Event Scale (INES). High unlikelihood of nuclear accidents at this level is demonstrated by the fact that there have been only two accidents at the maximum magnitude of seven (the nuclear accident at Chernobyl's reactor #4 in 1986 and the Fukushima #1 power plant in 2011) and one of magnitude six (the Kyshtym disaster in 1957). Many *environmental disasters* could also be considered a technology disaster, such as pollution, including atmospheric, water, sea, and soil pollution. Other environmental disasters, such as deforestation, may not be related to a technology issue but with other issues, such as economic development needs. Still, regardless of the type of cause, environmental disasters cannot be considered natural disasters because their origins and evolution are directly related to human actions.

References

[1] A. Kwasinski, J. Eidinger, A. Tang, and C. Tudo-Bornarel, "Performance of electric power systems in the 2010–2011 Christchurch New Zealand earthquake sequence." *Earthquake Spectra*, vol. 30, no. 1, pp. 205–230, Feb. 2014.

[2] Z. Abuza, *The Ongoing Insurgency in Southern Thailand: Trends in Violence, Counterinsurgency Operations, and the Impact of National Politics*. Institute for National Strategic Studies (INSS), Strategic Perspectives No. 6, National Defense University, Washington, DC, Sept. 2011.

[3] R. Lordan-Perret, A. L. Wright, P. Burgherr, M. Spada, and R. Rosner, "Attacks on energy infrastructure targeting democratic institutions." *Energy Policy*, vol. 132, pp. 915–927, Sept. 2019.

[4] A. Villemeur, *Reliability, Availability, Maintainability, and Safety Assessment*. Volume 1, Methods and Techniques, John Wiley and Sons, West Sussex, UK, 1992.

[5] A. Kwasinski, W. Weaver, and R. Balog, *Micro-grids in Local Area Power and Energy Systems*, Cambridge University Press, Cambridge, 2016.

[6] A. Kwasinski and P. T. Krein, "Optimal configuration analysis of a microgrid-based telecom power system," in Rec. INTELEC 2006, pp. 602–609.

[7] K. Yotsumoto, S. Muroyama, S. Matsumura, and H. Watanabe, "Design for a Highly Efficient Distributed Power Supply System Based on Reliability Analysis," in Proceedings of INTELEC 1988, pp. 545–550.

[8] IEEE Standards Association (IEEE SA). "IEEE Guide for Electric Power Distribution Reliability Indices," IEEE Std 1366–2003 (Revision of IEEE Std 1366–2003), 2004.

[9] M. D. Powell and T. A. Reinhold, "Tropical cyclone destructive potential by integrated kinetic energy." *Bulletin of the American Meteorological Society*, vol. 88, no. 4, pp. 513–526, Apr. 2007.
[10] V. Krishnamurthy and A. Kwasinski, "Characterization of Power System Outages Caused by Hurricanes through Localized Intensity Indices," in Proceedings of the 2013 IEEE Power and Energy Society General Meeting, pp. 1–5.
[11] G. Cruse and A. Kwasinski, "Statistical Evaluation of Flooding Impact on Power System Restoration Following a Hurricane," in Proceedings of 2021 Resilience Week, October 20, 2021.
[12] J. B. Wardman, T. M. Wilson, P. S. Bodger, J. W. Cole, and C. Stewart, "Potential impacts from tephra fall to electric power systems: a review and mitigation strategies." *Bulletin of Vulcanology*, vol. 74, pp. 2221–2241, Sept. 2012.
[13] N. U. Crooker, "Solar and heliospheric geoeffective disturbances." *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 62, no. 12, pp. 1071–1085, Dec. 2000.
[14] Australian Government, "Solar Coronal holes," www.ips.gov.au/Category/Educational/The%20Sun%20and%20Solar%20Activity/General%20Info/Solar_Coronal_Holes.pdf.
[15] V. D. Albertson, B. Bozoki, W. E. Feero et al., "Geomagnetic disturbance effects on power systems." *IEEE Transactions on Power Delivery*, vol. 8, no. 3, pp. 1206–1216, July 1993.
[16] L. Trichtchenko and D. H. Boteler, "Effects of Recent Geomagnetic Storms on Power Systems," in Proceedings of the International Symposium on Electromagnetic Compatibility and Electromagnetic Ecology, 2007, pp. 265–268.
[17] NOAA Space Prediction Center, "The K-Index." www.swpc.noaa.gov/info/Kindex.html.
[18] National Research Council, "Severe Space Weather Events: Understanding Societal and Economic Impacts Workshop Report," Committee on the Societal and Economic Impacts of Severe Space Weather Events: A Workshop, 2008.
[19] V. D. Albertson and J. M. Thorson, "Power System Disturbances during a K-8 Geomagnetic Storm: August 4, 1972." *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-93, no. 4, pp. 1025–1030, July 1974.
[20] J. Kolawole, S. Mulukulta, and D. Glover, "Effect of Geomagnetic-Induced-Current on Power Grids and Communication Systems: A Review," in Proceedings of Annual North American Power Symposium, 1990, pp. 251–262.
[21] J. G. Kappernman and V. D. Albertson, "Bracing for the geomagnetic storms." *IEEE Spectrum*, vol. 27, no. 3, pp. 27–33, Mar. 1990.
[22] V. D. Albertson, J. M. Thorson, and S. A. Miske, "The effects of geomagnetic storms on electrical power systems." *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-93 no. 4, pp. 1031–1044, July 1974.
[23] J. G. Kappenman, "Geomagnetic storms and their impact on power systems." *IEEE Power Engineering Review*, vol. 16, no. 5, pp. 5–8, May 1996.
[24] J. Kappenman, "Geomagnetic Storms and Their Impacts on the US Power Grid," Oak Ridge National Lab doc id Meta-R-319, Jan. 2010.
[25] P. R. Barnes, "Electric Utility Experience with Geomagnetic Disturbances," Oak Ridge National Lab doc id ORNL-6665, Sept. 1991.
[26] J. A. Marusek, "Solar Storm Threat Analysis," report from Impact Inc., 2007, www.breadandbutterscience.com/SSTA.pdf.
[27] J. G. Kappenman, W. A. Radasky, J. L. Gilbert, and L. A. Erinmez, "Advanced geomagnetic storm forecasting: a risk management tool for electric power system operations." *IEEE Transactions on Plasma Science*, vol. 28, no. 6, pp. 2114–2121, Dec. 2000.

[28] W.-M. Boerner, J. B. Cole, W. R. Goddard et al., "Impacts of solar and auroral storms on power line systems." *Space Science Reviews*, vol. 35, pp. 195–205, June 1983.
[29] J. L. Green and S. Boardsen, "Duration and extent of the Great Auroral Storm of 1859." *Advances in Space Research*, vol. 38, no. 2, pp. 130–135, 2006.
[30] S. F. Odenwald and J. L. Green, "Bracing the satellite infrastructure for a solar superstorm." *Scientific American*, July 2008.
[31] S. Glasstone and P. Dolan, "The Effects of Nuclear Weapons," United States Department of Defense and Department of Energy Technical Report, 1977.
[32] E. R. Fletcher, R. W. Albright, R. F. D. Perret et al., "Nuclear Bomb Effects Computer (Including Slide-rule Design and Curve Fits for Weapons Effects)," Civil Effects Test Operations, US Atomic Energy Commission, 1963.
[33] V. Krishnamurthy, B. Huang, A. Kwasinski, E. Pierce, and R. Baldick, "Generalized resilience models for power systems and dependent infrastructure during extreme events." *IET Smart Grid*, vol. 3, no. 2, pp. 194–206, Apr. 2020.
[34] V. Krishnamurthy and A. Kwasinski, "Modeling of distributed generators resilience considering lifeline dependencies during extreme events." *Risk Analysis*, vol. 39, no. 9, pp. 1997–2011, Sept. 2019.
[35] V. Krishnamurthy and A. Kwasinski, "Refueling Delay Models in Heterogenous Road Networks for Wireless Communications Base Station Gensets Operating in Extreme Conditions," in Proceedings of 2021 Resilience Week, October 20, 2021.
[36] V. Krishnamurthy and A. Kwasinski, "Modeling of Communication Systems Dependency on Electric Power during Nuclear Attacks," in Proceedings of IEEE INTELEC 2016, Austin, TX, Oct. 2016.
[37] M. Kaku and D. Axelrod, *To Win a Nuclear War: The Pentagon's Secret War Plans*. Black Rose Books Ltd., Quebec, Canada, 1987.
[38] A. Kwasinski, F. Andrade, M. J. Castro-Sitiriche, and E. O'Neill, "Hurricane Maria effects on Puerto Rico electric power infrastructure." *IEEE Power and Energy Technology Systems Journal*, vol. 6, no. 1, pp. 85–94, Mar. 2019.