17 A hierarchical security architecture for smart grid

Quanyan Zhu and Tamer Başar

17.1 Introduction

The smart grid aims to provide reliable, efficient, secure, and quality energy generation/distribution/consumption using modern information, communications, and electronics technologies. The integration with modern IT technology moves the power grid from an outdated, proprietary technology to more common ones such as personal computers, Microsoft Windows, TCP/IP/Ethernet, etc. It can provide the power grid with the capability of supporting two-way energy and information flow, isolate and restore power outages more quickly, facilitate the integration of renewable energy resources into the grid and empower the consumer with tools for optimizing energy consumption. However, in the meantime, it poses security challenges on power systems as the integration exposes the system to public networks.

Many power grid incidents in the past have been related to software vulnerabilities. In [1], it is reported that hackers have inserted software into the USA power grid, potentially allowing the grid to be disrupted at a later date from a remote location. As reported in [2], it is believed that an inappropriate software update has led to a recent emergency shutdown of a nuclear power plant in Georgia, which lasted for 48 hours. In [3], it has been reported that a computer worm, Stuxnet has been spread to target Siemens SCADA systems that are configured to control and monitor specific industrial processes. On November 29, 2010, Iran confirmed that its nuclear program had indeed been damaged by Stuxnet [4, 5]. The infestation by this worm may have damaged Iran’s nuclear facilities in Natanz and eventually delayed the start up of Iran’s nuclear power plant.

Modern power systems do not have built-in security functionalities and the security solutions in regular IT systems may not always apply to systems in critical infrastructures. This is because critical infrastructures have different goals and assumptions concerning what needs to be protected, and have specific applications that are not originally designed for a general IT environment. Hence, it is necessary to develop unique security solutions to fill the gap where IT solutions do not apply.

In this chapter, we describe a layered-architecture to address the security issues in power grids, which facilitates identifying research problems and challenges at
each layer and building models for designing security measures for control systems in critical infrastructures. We also emphasize a cross-layer viewpoint toward security issues in power grids in that each layer can have security dependence on the other layers. We need to understand the tradeoff between the information assurance and the physical layer system performance before designing defense strategies against potential cyber threats and attacks. As examples, we address three security issues of smart grid at different layers, namely, the resilient control design problem at the physical power plant, the data-routing problem at the network and communication layer, and the information security management at the application layers.

The concept of hierarchical structures has been adopted as solutions for the Internet and manufacturing operations. The well-known layered structure of OSI model for the Internet has influenced the integration between software and hardware [6]. The upper layers of the OSI model represent software that implements network services like encryption and management. The lower layers implement more primitive, hardware-oriented functions like routing, addressing and flow control. The layered structure introduces a practical framework for network technology development at individual layers and also allow cross-layer methods to investigate issues across these virtual boundaries between layers [8].

The integration between enterprise and industrial control systems is guided by ISA95 standards for information exchange between enterprise and manufacturing control activities and their supporting IT systems. It defines levels within a manufacturing operation based on the Purdue Reference Model (PRM) for Computer Integrated Manufacturing (CIM) [7]. PRM has formed the basis for ISA95 standards today, providing openness necessary to unify plant resource requirements.

The hierarchical model in this chapter extends the notion from OSI and PRM models and integrates them for smart grids. Our model is related to the notion of resilient control systems proposed in [10, 11, 12]. The goal of a resilient control system is to maintain state awareness and an accepted level of operational normalcy in response to disturbances, including threats of an unexpected and malicious nature [10]. To achieve resilience, the control system design is divided into four parallel areas: human systems, complex networks, cyber awareness and data fusion. The hierarchical perspective shares the similar divide-and-conquer philosophy but views the system differently in a hierarchically structured way.

The model is also related to the one in [18], where a novel framework of security solution for power-grid automation has been proposed. The integrated security framework has three layers, namely, power, automation and control, and security management. The automation and control system layer monitors and controls power transmission and distribution processes, while the security layer provides security features.
17.2 Hierarchical architecture

Smart grid comprises of physical power systems and cyber information systems. The integration of the physical systems with the cyber space allows new degrees of automation and human-machine interactions. The uncertainties and hostilities existing in the cyber environment have brought emerging concerns for modern power systems. It is of supreme importance to have a system that maintains state awareness and an accepted level of operational normalcy in response to disturbances, including threats of an unexpected and malicious nature [10, 13].

The physical systems of the power grid can be made to be resilient by incorporating features such as robustness and reliability [14], while the cyber components can be enhanced by many cyber-security measures to ensure dependability, security, and privacy. However, the integration of cyber and physical components does not necessarily ensure overall reliability, robustness, security and resilience of the power system. The interaction between the two environments can create new challenges in addition to the existing ones. To address these challenges, we first need to understand the architecture of smart grids. Smart grid can be hierarchically organized into six layers, namely, physical layer, control layer, data communication layer, network layer, supervisory layer and management layer. This hierarchical structure is depicted in Figure 17.1. The first two layers, physical layer and control layer, can be jointly seen as physical environment of the system. The data communication layer and network layer comprise the cyber environment of the power grid. The supervisory layer together with the management layer constitute the higher-level application layer where services and human-machine interactions take place.

A simplified structure of power grid is depicted in Figure 17.2. The power plant is at the physical layer and the communication network and security devices are at the network and communication layers. The controller interacts with the communication layer and the physical layer. An administrator is at the supervisory layer to monitor and control the network and the system. Security management is at the highest layer where security policies are made against potential threats from attackers. SCADA is the fundamental monitoring and control architecture at the control area level. The control center of all major U.S. utilities have implemented a supporting SCADA for processing data and coordinating commands to manage power generation and delivery within the EHV and HV (bulk) portion of their own electric power system [15].

The layered structure is also commonly seen in SCADA systems [19]. In large SCADA systems, there is usually a communication network connecting individual PLCs to the operator interface equipment at the central control room. There are communication networks used at lower level in the control-system architecture for communication between different PLCs in the same subsystems or facility, as well as for communication between field devices and individual PLCs. Figure 17.3
Chapter 17. A hierarchical security architecture for smart grid

The information structure of SCADA systems in today’s power grids is highly hierarchical. Each primary controller utilizes its own local measurement only, and each control area utilizes measurements in its own utility only and has
its own SCADA system. Protection mechanisms are preprogrammed to protect individual pieces of equipment and rarely require communications [16, 17].

To further describe the functions at each layer, we resort to Figure 17.4, which conceptually describes a smart grid system with a layering architecture. The lowest level is the physical layer where the physical/chemical processes we need to control or monitor reside. The control layer includes control devices that are encoded with control algorithms that have robust, reliable, secure, fault-tolerant features. The communication layer passes data between devices and different layers. The network layer includes the data packet routing and topological features of control systems. The supervisory layer offers human-machine interactions and capability of centralized decision-making. The management layer makes economic and high-level operational decisions.

In subsequent subsections, we identify problems and challenges at each layer and propose problems whose resolution requires a cross-layer viewpoint.

17.2.1 Physical layer

The physical layer comprises the physical plant to be controlled. It is often described by an ordinary differential equation (ODE) model from physical or chemical laws. It can also be described by difference equations, Markov models, or model-free statistics. We have the following challenges that pertain to the security and reliability of the physical infrastructure. First, it is important to find appropriate measures to protect the physical infrastructure against vandalism, environmental change, unexpected events, etc. Such measures often need a cost-and-benefit analysis involving the value assessment of a particular infrastructure. Second, it is also essential for engineers to build the physical systems with more dependable components and more reliable architecture. It brings the concern on
the physical maintenance of the control system infrastructures that demands a cross-layer decision-making between the management and physical layers [22].

17.2.2 Control layer

The control layer consists of multiple control components, including observers/sensors, intrusion-detection systems (IDSs), actuators and other intelligent control components. An observer has the sensing capability that collects data from the physical layer and may estimate the physical state of the current system. Sensors may need to have redundancies to ensure correct reading of the states. The sensor data can be fused locally or sent to the supervisor level for global fusion. A reliable architecture of sensor data fusion will be a critical concern. An IDS protects the physical layer as well as the communication layer by performing anomaly-based or signature-based intrusion detection. An anomaly-based ID is more common for physical layer whereas a signature-based ID is more common for the packets or traffic at the communication layer. If an intrusion or an anomaly occurs, an IDS raises an alert to the supervisor or work
hand-in-hand with built-in intrusion prevention systems (related to emergency responses, e.g., control reconfiguration) to take immediate action. There lies a fundamental trade-off between local decisions versus a centralized decision when intrusions are detected. A local decision, for example, made by a prevention system, can react in time to unanticipated events; however, it may incur a high packet drop rate if the local decision suffers high false negative rates due to incomplete information. Hence, it is an important architectural concern on whether the diagnosis and control modules need to operate locally with IDS or globally with a supervisor.

17.2.3 The communication layer

Communication layer is where we have a communication channel between control layer components or network-layer routers. The communication channel can take multiple forms: wireless, physical cable, blue-tooth, etc. The communication layer handles the data communication between devices or layers. It is an important vehicle that runs between different layers and devices. It can often be vulnerable to attacks such as jamming and eavesdropping. There are also privacy concerns of the data at this layer. Such problems have been studied within the context of wireless communication networks [20]. However, the goal of a critical infrastructure may distinguish themselves from the conventional studies of these issues.

17.2.4 Network layer

The network layer concerns the topology of the architecture. It comprises of two major components: one is network formation and the other one is routing. We can randomize the routes to disguise or confuse the attackers so as to achieve certain security or secrecy or minimum delay. Moreover, once a route is chosen, how much data should be sent on that route has long been a concern for researchers in communications [21, 23, 28]. In control systems, many specifics to the data form and rates may allow us to reconsider this problem in a control domain.

17.2.5 Supervisory layer

The supervisory layer coordinates all layers by designing and sending appropriate commands. It can be viewed as the brain of the system. Its main function is to perform critical data analysis or fusion to provide immediate and precise assessment of the situation. It is also a holistic policy maker that distributes resources in an efficient way. The resources include communication resources, maintenance budget as well as control efforts. In centralized control, we have one supervisory module that collects and stores all historical data and serves as a powerful data-fusion and signal-processing center [30, 31].
17.2.6 Management layer

The management layer is a higher-level decision-making engine, where the
decision-makers take an economic perspective toward the resource allocation
problems in control systems. At this layer, we deal with problems such as: (i)
how to budget resources to different systems to accomplish a goal; and (ii) how to
manage patches for control systems, e.g., disclosure of vulnerabilities to vendors,
development and release of patches [24].

17.3 Robust and resilient control

The layered architecture in Figure 17.4 can facilitate the understanding of the
cross-layer interactions between the physical world and the cyber world. In this
section, we aim to establish a framework for designing resilient controller for
the physical power systems. In Figure 17.6, we describe a hybrid system model
that interconnects the cyber and physical environments. We use $x(t)$ and $\theta(t)$ to
denote the continuous physical state and the discrete cyber state of the system,
which are governed by the laws $f$ and $\Lambda$, respectively. The physical state $x(t)$
is subject to disturbances $w$ and can be controlled by $u$. The cyber state $\theta(t)$
is controlled by the defense mechanism $l$ used by the network administrator as
well as the attacker’s action $a$.

We view resilient control as a cross-layer control design, which takes into
account the known range of unknown deterministic uncertainties at each state
as well as the random unanticipated events that trigger the transition from one
system state to another. Hence, it has the property of disturbance attenuation
or rejection to physical uncertainties as well as damage mitigation or resilience
to sudden cyber attacks. It would be possible to derive resilient control for the
closed-loop perfect-state measurement information structure in a general setting
with the transition law depending on the control action, which can further be
simplified to the special case of the linear quadratic problem.

The framework depicted in Figure 17.6 can be used to describe the voltage
regulation problem of a power generator subject to sudden faults or attacks. A
power system has multiple generators interconnected through a large dynamic
network. A common approach to designing control systems for generators is
to model the dynamics of a single generator and to approximate everything
else as an infinite bus, i.e., the voltage and the phase of the entire network are
not affected by the input power or field excitation of the generator. Shown in
Figure 17.8, a single generator is connected to an infinite bus through parallel
transmission lines. It is possible to design a stabilizing control, called the power
system stabilizer (PSS), to damp out the low-frequency oscillations for a single-
machine infinite-bus (SMIB) system [32, 33, 34]. A fault can occur as a result of
an unanticipated cyber attack. For example, an attacker can break into the IT
system and damage the circuit breakers in a power grid, leading to an operation
Figure 17.6: The interactions between the cyber and physical systems are captured by their dynamics governed by the transition law $\Lambda$ and the dynamical system $f$. The physical system state $x(t)$ is controlled by $u$ with the presence of disturbances and noises. The cyber state $\theta(t)$ is controlled by the defense mechanism $l$ used by the network administrator as well as the attacker’s action $a$.

Figure 17.7: A two-state operation model for a power system with one normal operation state ($\theta = 1$) and one post-attack state ($\theta = 2$); $\lambda_{ij}, i, j = 1, 2$, are the transition rates between the two states.

under a faulty state. It is important that one designs a controller to regulate the system to equilibrium as quickly as possible if such a failure occurs [35], and at the same time a defense mechanism to protect the systems from possible attacks. In Figure 17.7, we describe a two-state operation. One is under the normal state ($\theta = 1$) and the other is the post-attack state ($\theta = 2$). We now consider a specific mathematical formulation:

Let $\delta$ be the power angle, $\omega$ the relative speed, $P_e$ the active electrical power delivered by the generator; and let $u_f$ be the input of the amplifier of the generator, taken as the control variable. The system equations to model SMIB are
Chapter 17. A hierarchical security architecture for smart grid

Figure 17.8: An illustration of a synchronous machine infinite bus system. A cyber attack can lead to the breaker failure and result in power loss. A control scheme is needed to regulate the system to equilibrium as quickly as possible after the attack.

Table 17.1: Table of parameters

<table>
<thead>
<tr>
<th>Symbol and value</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_c = 1$</td>
<td>Gain of the excitation amplifier</td>
</tr>
<tr>
<td>$D = 5.0$</td>
<td>Per-unit damping constant</td>
</tr>
<tr>
<td>$H = 4.0$</td>
<td>Per-unit inertia constant</td>
</tr>
<tr>
<td>$\omega_0 = 100\pi$</td>
<td>Synchronous machine speed</td>
</tr>
<tr>
<td>$P_m = 0.9$</td>
<td>Mechanical input power</td>
</tr>
<tr>
<td>$T_{d0} = 6.9$</td>
<td>Direct-axis transient short circuit time constant</td>
</tr>
<tr>
<td>$V_s = 0.83 + j0.38$</td>
<td>Infinite bus voltage</td>
</tr>
<tr>
<td>$x_d = 1.863$</td>
<td>Direct-axis reactance of the generator</td>
</tr>
<tr>
<td>$x'_d = 0.257$</td>
<td>Direct-axis transient reactance of the generator</td>
</tr>
<tr>
<td>$x_T = 0.127$</td>
<td>Reactance of the transformer</td>
</tr>
<tr>
<td>$x_L = 0.4853$</td>
<td>Reactance of the transmission line</td>
</tr>
</tbody>
</table>

then:

$$\dot{\delta}(t) = \omega(t)$$

$$\dot{\omega}(t) = -\frac{D}{2H} \omega(t) + \frac{\omega_0}{2H} (P_m(t) - P_e(t))$$

$$\dot{P}_e(t) = \frac{1}{T_{d0}'} P_e(t) + \frac{1}{T_{d0}'} \left\{ \frac{V_s}{x_{ds}} \sin(\delta(t)) \left[ k_c u f + T_{d0}' (x_d - x'_d) \frac{V_s}{x_{ds}} \omega(t) \sin(\delta(t)) \right] + T_{d0}' \omega(t) \cot(\delta(t)) \right\} + w,$$

(17.1)

where $w$ is the disturbance, $T_{d0}' = \frac{x'_{ds}}{x_{ds}} T_{d0}$; $x_{ds} = x_T + x_L + x_d$; $x'_d = x_T + x_L + x'_d$; the main parameters listed in Table 17.1 and their values are chosen based on [33].
Under normal operation ($\theta = 1$), the control objective is to regulate the synchronous machine states ($\delta, \omega, P_e$) to the level of ($\delta_0, 0, P_m$); here one can linearize system model (17.1) at ($\delta_0, 0, P_m$). An unanticipated fault caused by a cyber attack can happen at the rate $\lambda_{12}$. When a circuit breaker is compromised, the total reactance on the transmission line will change accordingly.

The transition rates $\lambda_{ij}, i, j = 1, 2$, take the following parametrized form: $\lambda_{12} = p, \lambda_{11} = -p, \lambda_{21} = \lambda_{22} = 0$, where we have assumed that the operation after the attacker cannot be immediately recovered. At the cyber layer, the administrator can take two actions, i.e., to defend ($l_1 = D$) and not to defend ($l_2 = \text{ND}$). The attacker can also take two actions, i.e., to attack ($a_1 = A$) or not to ($a_2 = \text{NA}$). The parameter $p$ determines the transition law with respect to pure strategies and its values are tabulated as follows.

In Table 17.2, we have assumed a higher probability of transition to a failure state if the attacker launches an attack while the cyber system does not have proper measures to defend itself. On the other hand, the probability is lower if the cyber system can defend itself from attacks. In the above table, we have assumed a base transition rate 0.05 to denote the inherent reliability of the physical system without exogenous attacks.

Let $x := [x_1, x_2, x_3]' = [\delta, \omega, P_e]'$ be the aggregate state. The goal of robust control is to find an optimal state-feedback control $u(t) = \mu'(t, x, \theta)$ so that the infinite-horizon cost function $L$, defined below, is minimized for the worst-case disturbance $w(t)$ for achieving the best noise attenuation [61]:

$$L(x, u, w; \theta, t_0) = \mathbb{E} \int_{t_0}^{\infty} (|x(t)|^2_{Q^1} + |u(t)|^2_{R^1} - \gamma^2 |w(t)|^2) dt,$$

(17.2)

where $\gamma > \gamma^*$ is a noise-attenuation level greater than the best achievable attenuation level $\gamma^*$, $|\cdot|$ denotes the Euclidean norm with appropriate weighting, $A^i, B^i, D^i, Q^i, R^i, i = 1, 2$, are matrices of appropriate dimensions, whose entries are continuous functions of time $t$, with $Q^i(\cdot) \geq 0, R^i(\cdot) > 0$, which are actually taken to be constant matrices. One can choose these weighting matrices:

$$Q^1 = Q^2 = \begin{bmatrix} 1000 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 10 \end{bmatrix}, \quad R^1 = 10, R^2 = 1,$$

which emphasize the regulation of power angle and the willingness to use more control under a post-attack state.

### Table 17.2: Defender vs. attacker

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>NA</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>0.1</td>
<td>0.05</td>
</tr>
<tr>
<td>ND</td>
<td>0.95</td>
<td>0.05</td>
</tr>
</tbody>
</table>
The optimal control $\mu^*$ depends on the interactions between the administrator and the attacker. Let $f$ and $g$ be the cyber strategies of the administrator and the attacker, respectively. The value function $V^i$, $i = 1, 2$, achieved under the optimal control $\mu^*$ is hence dependent on $f$ and $g$.

The goal of resilient control is to find an optimal cyber-policy $f^*$ to defend against attacker's optimal strategy $g^*$ at state $i = 1$. The defense against attacks happens on a longer time scale and involves decision-making at the human and cyber levels of the system. Using time-scale separation, the optimal defense mechanism can be designed by viewing the physical control system at its steady state at each cyber state $\theta$ at a given time $k$. The interaction between an attacker and a defending administrator can be captured by a zero-sum stochastic game with the defender aiming to maximize the long-term system performance or payoff function $V_\beta$, defined below, whereas the attacker aiming to minimize it:

$$V_\beta(i, f(k), g(k)) := \int_0^\infty e^{-\beta k} \mathbb{E}_i[f(k), g(k)] V^i(k, f(k), g(k)) dk,$$

where $\beta$ is the discount factor. The operator $\mathbb{E}_i$, $i = 1, 2$, is the expectation operator and $V^i(k, f, g)$ is the optimal value function at state $i$ with starting time at $k$, i.e.,

$$V^i(k, f, g) = \min_u \max_w L(x, u, w; i, k).$$  \hspace{1cm} (17.3)

We can consider a class of mixed stationary strategies $f$ that randomizes between $l_1 = D$ and $l_2 = ND$, and $g$ that randomizes between $a_1 = A$ and $a_2 = NA$. It can be seen that the problem of finding $f^*$ and $g^*$ depends on the optimal control $\mu^*$ at the physical layer. Hence, in order to find the optimal control $\mu^*$ and optimal defense $F^*$, one needs to solve the optimization problems (17.2) and (17.3) jointly.

Using jointly the optimality criteria for (17.2) and (17.3) \cite{61}, \cite{62}, one can obtain stationary saddle-point strategy $f^* = [1, 0]$, $g^* = [0, 1]$ and

$$\mu^*(t, x, 1) = -(R^1)^{-1} B^1 Z^1 x, \quad \mu^*(t, x, 2) = -(R^2)^{-1} B^2 Z^2 x,$$

where

$$Z^1 = \begin{bmatrix} 399.3266 & 31.8581 & -162.2334 \\ 31.8581 & 5.7083 & -15.2963 \\ -162.2334 & -15.2963 & 149.7459 \end{bmatrix},$$

and

$$Z^2 = \begin{bmatrix} 2.8512 & 0.1066 & -2.8575 \\ 0.1066 & 0.0345 & -0.1041 \\ -2.8575 & -0.1041 & 4.1506 \end{bmatrix}.$$

The saddle-point strategy $f^*$ and $g^*$ indicates that the defender should always be defending and the attacker has no intent to launch an attack. The attenuation level under $\mu^*$ for the worst-case disturbance is $\gamma^* = 8.5$. Figure 17.9 shows the
evolution of state $\omega$ (or $x_2$) with failure occurring at time $t = 10$. The optimal control design allows the state $\omega$ to be stabilized after fault occurs.

### 17.4 Secure network routing

One of the challenging issues at the data communication and networking layers of the smart grid is the assurance of secure routing of phasor measurement unit (PMU) and smart meter (SM) data in the open network, which is enabled by the adoption of IP-based network technologies. It is forecasted that 276 million smart grid communication nodes will be shipped worldwide during the period from 2010 to 2016, with annual shipments increasing dramatically from 15 million in 2009 to 55 million by 2016 [36]. The current dedicated network or leased-line communication methods are not cost-effective to connect large numbers of PMUs and SMs. Thus, it is foreseen that IP-based network technologies are widely adopted since they enable data to be exchanged in a routable fashion over an open network, such as the Internet [37, 38, 39, 40, 41]. This will bring benefits such as efficiency and reliability, and risks of cyber attacks as well. Without doubt, smart grid applications based on PMUs and SMs will change the current fundamental architecture of communication network of power grid, and bring new requirements for communication security. Delay, incompleteness and loss of PMU and SM data will adversely impact smart grid operation in terms of efficiency and reliability. Therefore, it is important to guarantee integrity and availability of those PMUs and SMs data. To meet the QoS requirements in terms of delay, bandwidth, and packet loss rate, QoS-based routing technolo-
gies have been studied in both academia and the telecommunications industry [42, 43, 44, 45]. Unlike video and voice, data communications of PMUs and SMs have different meanings of real-time and security, especially in terms of timely availability [37, 46, 47, 48, 49, 50]. Therefore, QoS-based and security-based routing schemes for smart grid communications should be studied and developed to meet smart grid application requirements in terms of delay, bandwidth, packet loss, and data integrity.

In this section, we leverage the hierarchical structure of power grids and propose a routing protocol that maximizes the QoS along the routing path to the control room. In addition, we optimize the data communication rates between the super data concentrator at the penultimate level with the control center. We propose a hybrid structure of routing architecture to enable the resilience, robustness and efficiency of the smart grid.

17.4.1 Hierarchical routing

Smart grid has a hierarchical structure that is built upon the current hierarchical power grid architecture. The end-users, such as households, communicate their power usage and pricing data with a local area substation which collects and processes data from SMs and PMUs. In the smart grid, the path for the measurement data may not be predetermined. The data can be relayed from smaller scale data concentrators (DCs) to some super data concentrators (SDCs) and then to the control room. With the widely adopted IP-based network technologies, the communications between households and DCs can be in a multi-hop fashion through routers and relay devices. The goal of each household is to find a path with minimum delay and maximum security to reach DCs and then substations. This optimal decision can be enabled by the automated energy management systems built in SMs. Figure 17.10 illustrates the physical structure of the smart grid communication network. The PMUs and SMs send data to DCs through a public network. DCs process the collected data and send the processed data to SDCs through (possibly) another public network.

The physical communication structure from local meters to the control room can be logically divided into several levels. Figure 17.11 illustrates a snapshot of routing paths in a simplified example of the hierarchical structure of the physical communication architecture, which is divided logically into 5 levels. For simplicity, we depict only level of routers in Figure 17.11. In practice, there can be multiple levels of routers and they can be found in the public network between DCs and SDCs as well as SDCs and the control room. In the depicted smart grid, the data from a PMU or a SM has to make four hops to reach the control room. The decision for a meter to choose a router depends on the communication delay, security enhancement level and packet loss rate. In addition, the decisions for a DC to choose a SDC also depends on the same criteria. The communication security at a node is measured by the number of security devices such as firewalls, intrusion-detection systems (IDSs) and intrusion-prevention systems.
A hierarchical security architecture for smart grid

Figure 17.10: An example of the physical structure of the hierarchical smart grid communication network.

Figure 17.11: A snapshot of routing paths in the hierarchical smart grid.

(IPSs) deployed to reinforce the security level at that node. We assign higher utility to network routers and DCs that are protected by a larger number of firewalls, IDSs/IPSs and dedicated private networks in contrast to public networks. This relatively simple metric only considers one aspect of the control system cyber security. It can be further extended to include more security aspects by considering the authorization mechanisms, the number of exploitable vulnerabilities, potential damages as well as recovery time after successful attacks. The readers can refer to [51, 52, 13, 53] for more comprehensive metrics.
A trade-off with higher security is the latency and packet loss rate incurred in data transmission. A secure network inevitably incurs delays in terms of processing (encrypting/decrypting) and examining data packets. We can model the process of security inspection by a tandem queueing network. Each security device corresponds to an M/M/1 queue whose external arrival rate follows a Poisson process and the service time follows an exponential distribution. The latency caused by the security devices such as IDSs/IPSs is due to the number of predefined attack signatures and patterns to be examined [22, 26, 29]. In addition, devices such as IPSs can also lead to high packet loss due to their false negative rates in the detection.

Furthermore, a node with higher level of security may be preferred by many meters or routers, eventually leading to a high volume of received data and hence higher level of congestion delay. This fact enables a game-theoretical approach to analyze the distributed routing decisions in the smart grid. The solution concept of mixed Nash equilibrium [9] as a solution outcome is desirable for two reasons. First, in theory, mixed Nash equilibrium always exists for a finite matrix game [9] and many learning algorithms such as fictitious play and replicator dynamics can lead to mixed Nash equilibrium [54, 25, 27]. Second, the randomness in the choice of routes makes it harder for an attacker to map out the routes in the smart grid.

17.4.2 Centralized vs. decentralized architectures

Section 17.4.1 introduced a hierarchical routing problem for the smart grid. The decision at the penultimate level is trivial since all the data should route to the control station. In reality, a local control station pulls data from its communicating SDCs. SDCs cannot send data to the control station at an arbitrary rate. The bandwidth and communication resources of a control station are often constrained. Hence, it becomes important to consider an appropriate resource allocation at the data communication between SDCs and the control station. The resource allocation scheme can be either centralized, i.e., determined by the control station, or decentralized, i.e., determined by individual SDCs.

A centralized routing architecture ensures the global efficiency and it is robust to small disturbances from SMs and individual DCs or SDCs. However, it is costly to implement centralized planning on a daily basis for a large-scale smart grid. In addition, global solutions can be less resilient to unexpected failures and attacks as they are less nimble for changes in routes and it takes time for the centralized planner to respond in a timely manner.

On the other hand, decentralized decision-making can be more computation-ally friendly based on local information and hence the response time to emergency is relatively fast. The entire system becomes more resilient to local faults and failures, thanks to the independence of the players and the reduced overhead on the response to unanticipated uncertainties. However, the decentralized solution can suffer from high loss due to inefficiency [30, 31]. Hence, we need to
assess the tradeoff between efficiency, reliability and resilience for designing the communication protocol between the control stations and the SDCs.

Illustrated in Figure 17.12, a hybrid architecture of the communication infrastructure of $L$ levels allows to incorporate desirable features of the two architectures. One can adopt a centralized planning at the top levels $L - 1$ and $L$ while building a decentralized routing protocol between levels 1 and $L - 1$. Such an architecture can enable robustness at the critical data centers and resilience at the lower user level.

Hence, the last-stage utility is determined in a centralized manner by the control room based on the priority and load. The resource allocation at the last level is robust to small parametric disturbances and is independent of routing decisions between level 1 and level $L - 1$. The routing decisions of the meters are resilient to router failures in a public network. The learning algorithms in [54, 25, 27] can be used for responding to a dysfunctional router by selecting a new router once the one in use is compromised.

17.5 Management of information security

The use of technologies with known vulnerabilities exposes power systems to potential exploits. In this section, we discuss information security management which is a crucial issue for power systems at the management layer in Figure 17.4. The timing between the discovery of new vulnerabilities and their patch availabilities is crucial for the assessment of the security risk exposure of software users [55, 57, 58, 59]. The security focus in power systems is different from the one
in computer or communication networks. The application of patches for control systems needs to take into account the system functionality, avoiding the loss of service due to unexpected interruptions. The disclosure of software vulnerabilities for control systems is also a critical responsibility. Disclosure policy indirectly affects the speed and quality of the patch development. Government agencies such as CERT/CC (Computer Emergency Response Team/Coordination Center) currently act as a third party in the public interest to set an optimal disclosure policy to influence behavior of vendors [60].

The decisions involving vulnerability disclosure, patch development and patching are intricately interdependent. In Figure 17.13, we illustrate the relationship between these decision processes. A control system vulnerability starts with its discovery. It can be discovered by multiple parties, for example, individual users, government agencies, software vendors or attackers and hence can incur different responses. The discoverer may choose not to disclose it to anyone, may choose to fully disclose through a forum such as bugtraq [56], may report to the vendor, or may provide to an attacker. Vulnerability disclosure is a decision process that can be initiated by those who have discovered the vulnerability. Patch development starts when the disclosure process reaches the vendor and finally a control system user decides on the application of the patches once they become available. An attacker can launch a successful attack once it acquires the knowledge of vulnerability before a control system patches its corresponding vulnerabilities. The entire process illustrated in Figure 17.13 involves many agents or players, for example, system users, software vendors, government agencies, attackers. Their state of knowledge has a direct impact on the state of vulnerability management.

We can compartmentalize the task of vulnerability management into different submodules: discovery, disclosure, development and patching. The last two submodules are relatively convenient to deal with since the agents involved in the decision making are very specific to the process. The models for discovery
and disclosure can be more intricate in that these processes can be performed by many agents and hence specific models should be used for different agents to capture their incentives, utility, resources and budgets.

17.5.1 User patching

In this subsection, we establish a model for users to determine the optimal time to patch their control systems. In control systems, it is known that the attack rate is low and the patching rate is low as well. It often occurs that users do not patch until there is a security alert, an available patch announcement or an experienced security breach. The operation of control system is separated into several operating periods. A control system cannot halt its operation until the end of an operating period. Let \( B_k, k = 1, 2, \ldots \), denote the \( k \)-th operating period since the last patching and \( T_k = T_{k-1} + B_k = \sum_{i=1}^{k} B_i \), where \( T_0 \) is the beginning of the first operation period. Let \( \tau \) be the time length between the start of the first operation period and a security alert or an attack. Let \( f_\tau(t), t \geq 0 \), be its probability density function, which is taken to be hyper-exponential with \( n \) phases and parameters \( \lambda = (\lambda_1, \ldots, \lambda_n) \) and normalized weighting factor \( q = (q_1, \cdots, q_n) \), i.e.,

\[
f_\tau(t) = \sum_{i=1}^{n} q_i g_i(t) = \sum_{i=1}^{n} q_i \lambda_i \exp(-\lambda_i t), \quad \sum_{i=1}^{n} q_i = 1. \quad (17.4)
\]

Each phase \( i \) can be interpreted differently. For example, let \( g_1(t) \) be the distribution of the arrival rate of alerts; \( g_2(t) \) be the arrival rate of an attack; \( g_3(t) \) be the arrival rate of an announcement of an available patch. We illustrate this model in Figure 17.14. At every \( T_{k-1}, k \geq 1 \), a control system starts to operate for a period of \( B_k \) and then stops for monitoring and patching.

The decision of an administrator is to determine \( B_k \) so that the risk of an unpatched control system subject to potential attacks is minimized. The decision-making process can be viewed as a black box as in Figure 17.15, where the decision input is the knowledge of the arrival rates of an attack; the intrinsic system parameters are the monitoring cost \( c_m \) and the production cost parameters \( c_0 \) and \( c_1 \); and the decision output is the operation period \( B_k \). The input and output
characteristics of the decision process assist us in integrating it into the system model in Figure 17.13.

Let $c_m \in \mathbb{R}_+$ be the monitoring cost at the end of each operation period and $c_p(t_k, b_{k+1}) : \mathbb{R}_+^2 \to \mathbb{R}_+$ be the operation cost of running the plant for $b_{k+1}$ starting from $t_k$. We have the following dynamic programming (DP) equation to find the optimal decision policy to be taken at each period, taking into account the whole lifetime of the system:

\[
V^*_k(t_k) = \min_{b_{k+1} \geq 0} \left\{ E[c(t_k, b_{k+1}) + P(\tau_{t_k} > b_{k+1})V^*_{k+1}(t_{k+1})] \right\}, \tag{17.5}
\]

where $b_k$ is the decision variable of operating time; $V^*_k(t_k)$ represents the optimal cost at time $t_k$. The term $P(\tau_{t_k} > b_{k+1})$ represents the transition probability given by the conditional probability

\[
P(\tau_{t_k} > b_{k+1}) := P(\tau > t_k + b_{k+1}|\tau > t_k). \tag{17.6}
\]

The term $E[c(t_k, b_{k+1})]$ is the stage cost at $t_k$ given by

\[
E[c(t_k, b_{k+1})] = \hat{\gamma}E[(b_{k+1} - \tau_{t_k})1(\tau_{t_k} \leq b_{k+1})] + c_m + c_p(t_k, b_{k+1}), \tag{17.7}
\]

where $\hat{\gamma}$ denotes the unit cost of untimely patching; $\tau_{t_k}$ is the conditional residual time counting from $t_k$ given that $\tau_{t_k} > t_k$. By solving the DP equation, we can find a dynamic policy for the operation of the plants at each starting time $t_k$.

We can simplify the general model by assuming that the arrival of security alerts or breaches form a single Poisson process with rate $\lambda$ and the arrival time $\tau$ and the conditional residual time $\tau_{t_k}$ are exponentially distributed with parameter $\lambda$. Let $C^0_k = \frac{1}{2}c_0 b_k^2$ be the cost for operating a plant non-stop for a period of time $b_k$. Denote by $C^1_k = c_1 b_k$ the linear gain or profit from running the plant. Hence, we can assume that the cost $c_p$ is given by

\[
c_p(t_k, b_{k+1}) := C^0_k - C^1_k = \frac{1}{2}c_0 b^2_k - c_1 b_k. \tag{17.8}
\]

Due to the memoryless property of exponential distribution, the DP equation in (17.5) can be simplified to

\[
V^*(\lambda) = \min_{b \geq 0} \left\{ \hat{\gamma}E[(b - \tau(\lambda))1(\tau(\lambda) \leq b)] + \frac{1}{2}c_0 b^2 - c_1 b + P(\tau(\lambda) > b)V^*(\lambda) \right\}.
\]
For each fixed $b \geq 0$, $V(\lambda)$ can be solved from (6) (without the minimum) to yield

$$V(\lambda, b) = \frac{\bar{c} \mathbb{E}[(b - \tau(\lambda))1_{(\tau(\lambda) \leq b)}] + \frac{1}{2}c_0b^2 - c_1b}{1 - P(\tau(\lambda) > b)}. \quad (17.9)$$

Note that in the above,

$$\mathbb{E}[(b - \tau(\lambda))1_{(\tau(\lambda) \leq b)}] = \frac{\lambda b - 1 + \exp(-\lambda b)}{\lambda}, \quad (17.10)$$

and the term in the denominator of (17.9) is as follows:

$$P(\tau(\lambda) > b) = \exp(-\lambda b). \quad (17.11)$$

Hence, solving the DP is equivalent to finding a solution to the optimization problem (R-OPT) that takes into account the risk factor of potential threats and attacks:

$$\text{(R-OPT)} \quad V^*(\lambda) = \min_{b \geq 0} V(\lambda, b). \quad (17.12)$$

Note that a simple solution for operation time $b$ without security risk consideration, i.e., ignoring the potential costs that can be incurred by attacks in (R-OPT), is based on a cost and benefit analysis solving the risk-free optimization problem (NR-OPT):

$$\text{(NR-OPT)} \quad \min \frac{1}{2}c_0b_k^2 - c_1b_k + c_m, \quad (17.13)$$

which yields a benchmark solution $b^* = c_1/c_0$.

We can numerically solve the DP equations for a given scenario of control systems. Set the parameters $c_1 = 10$, $c_0 = 0.1$, $\gamma = 10$, $c_m = 10$, $\lambda = 0.001$ as the nominal case. In Figure 17.16, we show the optimal operation period versus the monitoring cost. It can be seen that when the cost becomes higher, the control system cannot afford a frequent checking and monitoring and hence the operation period increases. In Figure 17.17, we show the optimal operation period versus the attack rate. We observe that when the attack rate is high, the control system need to decrease its operation period to monitor and update its system more often. From the simulation results, we notice that an optimal operation period of control systems given the currently estimated attack rate is roughly around half a month.

### 17.6 Conclusion

Security issues that arise in the smart grid constitute a pivotal concern in modern power-system infrastructures. In this chapter, we have discussed a six-layer secu-

---

1 “R-OPT” stands for risk-based optimization.
2 “NR-OPT” stands for no-risk-based optimization.
Chapter 17. A hierarchical security architecture for smart grid

Figure 17.16: Operation period vs. monitoring cost.

Figure 17.17: Operation period vs. attack rate.

security architecture for the smart grid, motivated by the OSI for the Internet and PRM models for enterprise and control systems. We have identified the security challenges present at each layer and pinpointed a holistic viewpoint for security solutions in the smart grid. The layered architecture facilitates the understanding of the tradeoff between the information assurance at the cyber-related layers and the physical layer system performance.

We have addressed security issues at three different layers. The resilient control design at the physical system is pivotal for modern power systems. We have discussed a hybrid framework in which the occurrence of unanticipated events is modeled by a stochastic switching, and deterministic uncertainties are represented by the known range of disturbances. This framework has taken resilience of physical systems into consideration and has enabled a cross-layer control design for modern power grids.

At the data communication and network layers, we have investigated the secure routing problem in the smart grid, which arises from the adoption of IP-based network technologies due to the wide use of PMUs and smart meters. We have leveraged the hierarchical structure of power grids and discussed a routing proto-
col that is based on distributed optimization of the quality-of-service along individual routing paths. We have also illustrated the hybrid structure of the routing protocol to incorporate the desirable features of the centralized and decentralized architectures.

The use of information technologies in power systems poses additional potential threats due to the frequent disclosure of software vulnerabilities. At the higher level of the information security management layer, we have discussed a series of policy-making decisions on vulnerability discovery, disclosure, patch development and patching. We have adopted a system approach to understand the interdependencies of these decision processes.

The future work that follows the hierarchical model involves the integration of learning algorithms for detection, automation and reconfiguration in the smart grid. In addition to the security problems illustrated in the chapter, there are other security and privacy issues existing at each layer, for example, the jamming and eavesdropping problems at the data communication layer, the user data privacy problem at the management layer, and the system reliability problem at the network layer. Furthermore, the hierarchical framework can be extended to study multi-agent systems. The interactions between sub-systems in the smart grid can reside at the network, communication and physical layers. It will be interesting to investigate the competition and cooperation for resources at multiple layers.
References


[26] Q. Zhu and T. Başar, “Dynamic policy-based IDS configuration,” in Proc. of 48th IEEE Conference on Decision and Control (CDC), Shanghai, P. R.
Chapter 17. A hierarchical security architecture for smart grid


The new era of vulnerability disclosure—a brief chat with HD Moore


