CYPRESS

Arbitrating between preventive and corrective cyber-physical risk mitigation

Task 3.1 - Project Report

Efthymios Karangelos, Amirreza Jafari Anarjan, Louis Wehenkel

Date: November 2023

Contents

Executive Summary

The CYPRESS project aims at developing novel knowledge, methods and tools needed to help ensure the security of supply through the transmission grid, while accounting for the specific nature of cyber-threats and integrating them into a coherent probabilistic risk management approach. It is articulated along three research themes, aiming to develop: i) novel models and benchmarks for computer simulation and laboratory testing of the cyber-physical electric power system security of supply, ii) techniques for assessing the cyber-physical security of electric energy supply, and iii) techniques for enhancing the cyber-physical security of electric energy supply. The project scope falls entirely within the category of "fundamental research" within the meaning of Regulation (EU) No 651/2014 because it is experimental and theoretical work undertaken essentially with a view to acquire new knowledge on the foundations of phenomena or observable facts. The project is not intended to develop commercial tools.

The work presented in this document has been performed in the frame of CYPRESS WP3, titled "Mitigation of cyber-physical security risks". The objective of CYPRESS WP3 is to develop methods and algorithms to help reducing the risk with respect to the cyber-physical vulnerabilities of the electric power system. The document is the outcome of task 3.1, titled "Arbitrating between preventive and corrective cyber-physical risk mitigation". The objective of this task was to extend multi-stage stochastic programming approaches that have already been proposed to arbitrate between preventive and corrective measures in the context of physical power system security, so as to cover cyber-threats and in particular malicious attacks. This includes investigating the possible mitigation measures that could be applied in preventive and/or in corrective mode and in fine proposing optimization problem formulations that allow one to arbitrate among them in a well-informed way.

Following an introductory chapter, the 2nd chapter of this report provides a brief overview to the application of multi-stage stochastic programming in the context of electric power system (physical) security management. The purpose is to establish the necessary background for the extension of such approaches towards cyber-physical security management. This is done by stating model-agnostic stochastic formulations for the interrelated problems of power systems real-time operation and operation planning. Next, chapter 3 presents an investigation on both the cyber-physical threats facing the electric power system as well as the cyber countermeasures that may be deployed in preventive and/or corrective mode to counteract these. A review of the literature indicates that, up to nowadays, the focus of the research effort has been on single, precisely defined, threat instances rather than the (more general) problem of identifying the suite of countemeasures that should be put in place to sufficiently protect the system against the broad spectrum of unknown threats it faces.

Chapter 4 formalizes the extension of the multi-stage stochastic programming approach from the domain of physical security to the domain of cyber-physical security management. A main challenge

for this is the multitude of complex functionalities of the power grid cyber sub-system, in turn translating into a multitude of cyber-physical threats with diverse modeling requirements. We have inevitably opted for generality. More specifically, we first introduce a modeling abstraction of the power grid cyber sub-system as an interface between the physical procesess of electricity generation, transmission and distribution and the power system operator. We next use this modeling abstraction to state model-agnostic formulations for the decision making problem of malicious cyber-physical attackers[1]. We finally take a step back in time and discuss alternative generic formulations for the decision making problem of a so-called cyber-physical operation planner (a.k.a. security manager). This actor is seeking to identify optimal preventive/corrective cyber and physical security measures and while facing uncertainty on the properties of the malicious actor threatening the power grid.

The next steps in this research effort are discussed in chapter 5. The continuation of the CYPRESS WP3 research effort concerns both the precise mathematical models that should be used to formulate relevant instances of these stochastic problems as well as the development of proof-of-concept solution approaches.

Author contributions

Efthymios Karangelos is the author of chapters 1, 2,4,5 and co-author of chapter 3 and appendix A. Amirreza Jafari Anarjan is a co-author of chapter 3 and editor of the report. Louis Wehenkel is a co-author of appendix A and editor of the report.

Author	Affiliation
Efthymios Karangelos (Task leader)	Université de Liège
Amirreza Jafari Anarjan	Katholieke Universiteit Leuven
Louis Wehenkel	Université de Liège

Table 1: List of Authors

[1] An instance of such model-agnostic formulation, specifying precise models and data as well as the strategy and motivation of a cyber-physical attacker is given in Appendix A.

1

Introduction

The continuous availability of electricity supply is of paramount importance to modern society. The practically uninterrupted service that the power grid end-users consider as a granted is certainly not an outcome of luck. Keeping this extra large and extremely complex system in constant operation entails a very challenging series of interrelated decisions and respective actions. Electric power system *security management* is the process of taking such decisions. Its overall purpose is to ensure in advance that the power grid will continue to operate, in spite of its vulnerabilities and the threats it may face in the future. Formally the aim of security management is to achieve a *security criterion* which concretely defines the sought level of assurance in terms of the ability of the system to continue operating even upon occurrence of a specific (sub-)set of credible threats.

Up to nowadays, the focus of security management has been on the physical vulnerabilities and threats facing the electric power grid. The so-called *N-1 security criterion* remains the central concept in today's practice. It prescribes that the bulk power system should continue to operate even after the forced outage of any single component. Doing so requires placing barriers in advance to ensure that the system can safely "ride-through" a threat, and/or promptly reacting to a threat before it compromises the electricity supply service to the system end-users. In power engineering terminology, the former is called *preventive* control while the latter is referred to as *corrective* control. Even with the well understood N-1 criterion, the choice between preventive and corrective control is not trivial. Preventive control is applied (and payed for) in advance, even if a threat may not materialize. Corrective control has to be applied in a prompt manner upon the occurrence of a threat, under stressful and dynamically evolving conditions [1]. Multi-stage stochastic programming offers a systematic solution to the dilemma, by mathematically formalizing the process of choosing the best strategy on the basis of physical models for the system and of the statistical properties of the concerned threats. More recently, it is also used as the primary tool to address pertinent planning and operation questions concerning the growing penetration of uncertain renewable power generation resources.

The topic of this report is the extension of the current multi-stage stochastic programming ap-

proaches from the domain of physical security management to the domain of cyber-physical security management. Such extension is certainly not straightforward. Cyber-physical threats are not purely fatal and they do not necessarily show the "favorable" statistical properties of physical threats. Rather, this class of threats even includes the behavior of malicious adversaries that intentionally seek to exploit a combination of cyber sub-system and power grid vulnerabilities to maximize disruption. In doing so, malicious adversaries do not only restrict to known system vulnerabilities but also actively try to identify (and even possibly create) unknown ones. As if these features are not complicating enough, cyber-physical security management requires modeling both the cyber sub-system and the physical power grid as well as their interrelation.

In light of the overwhelming complexity, most research efforts to date have focused on decision making problems related to a specific single threat (*i.e.*, the threat posed by a specific cyber-attacker with a precise objective and access to specific parts of the power grid cyber sub-system at certain moment in time)[1]. Our goal is to formalize a general decision making approach that does not only focus on a single specific instance and a precisely defined cyber-physical threat. Indeed, we consider that a cyber-physical security management approach should simultaneously protect the system from a broad spectrum of alternative eventualities rather than a single specific threat (in the same way that N-1 security management protects the grid from a set of alternative contingency events). These alternative eventualities may be used to describe the uncertainty of the decision maker potentially facing an external threat.

To do so, we start in chapter 2 by revisiting the current applications of stochastic programming approach for security management at the physical layer of the electric power system. Next, chapter 3 focuses on the alternative classes of cyber-physical threats as well as on the countermeasures acting on the cyber sub-system. Chapter 4 is dedicated on the development of a decision making framework for cyber-physical power system security management. We introduce alternative formulations for making security management choices on the cyber sub-system and on the power grid and discuss the associated computational feasibility and complexity. Chapter 5 concludes by discussing the stakes for probabilistic cyber-physical power system security management and the next steps of this research effort.

[1] As evidenced in multiple literature surveys [2]–[6].

2

Stochastic programming for power system physical security management

This chapter presents a synthetic overview of the state-of-the-art stochastic programming approaches for bulk electric power system security management. Referring the reader back to the 1st report of the CYPRESS project [7] for the fundamentals of power system security management we focus here on introducing the mathematical statement of the corresponding decision-making problems under uncertainty.

2.1 Temporal decomposition

Ensuring the continuous supply of electricity involves a series of decisions taken over different temporal horizons and facing distinct uncertainties. Decisions to build new infrastructure are taken over an horizon of (several) decade(s) and while also facing so-called *macro* uncertainties (*e.g.*, the political mood, the evolution of climate change). Operational decisions in contrast are taken with an horizon of minutes to days, and while only facing the *micro* uncertainties of daily life (*e.g.*, the unpredictability of renewable power generation). The series of decisions is interrelated in the sense that anterior actions (*e.g.* building a new transmission line) should anticipate on future decisions (the way that congestion will be managed in the future) while posterior actions are restricted by all previous decisions that have been irreversibly enacted on the power system. Taking all such interrelated decisions at once, by solving a single integrated stochastic program is obviously impractical. Rather, as per the so-called *"Russian dolls approach"*, simplified models of shorter-term decision-making problems are nested within longer-term problems, Fig. 2.1.

In the operational timescales, the Russian doll approach translates into the decomposition between the so-called *Operation Planning* and *Real-time Operation* stochastic decision-making problems.

Figure 2.1: The Russian dolls

The latter is the most well-studied problem of taking decisions to ensure that the system can withstand component outages over the next few minutes to hours. It covers the bottom three layers of Fig. 2.2 and is mathematically formalized in section 2.2. The Operation planning problem also includes the top layer of Fig. 2.2, seeking decisions that will enable real-time operation over the next hours to days. It can be understood as a *holy grail* for current research & development due to the proliferation of renewable generation and the associated daily uncertainty. Section 2.3 is dedicated to introducing the most prominent alternative approaches for this problem.

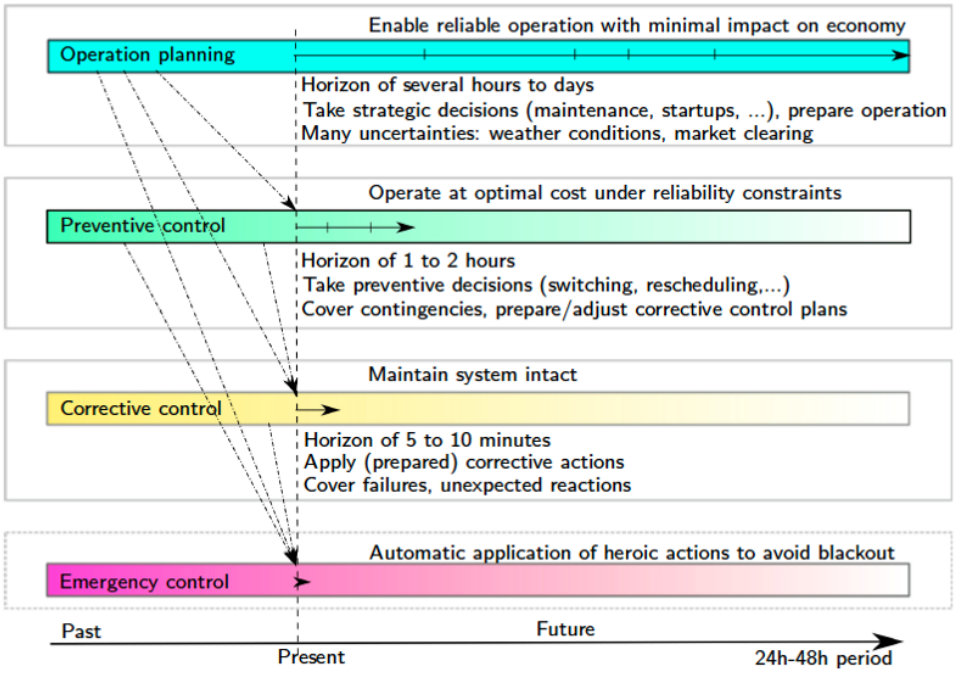

Figure 2.2: Decomposition of operational problems (credit: L. Wehenkel)

2.2 Real-time operation

In the current approach for physical power system security, the notion of preventive *vs* corrective control is mostly associated with credible contingencies (*i.e.*, unanticipated forced outages of bulk power system components). Referring the reader to [8], [9] for a detailed presentation of the state-of-the-art, let us introduce here the main variables, parameters, constraints and objective(s) of the so-called *Security Constrained Optimal Power Flow* (SCOPF) problem, which is compactly formulated as follows:

$$\min_{\mathbf{x},\mathbf{u}} \left\{ f_0(\mathbf{u}_0, \mathbf{y}_0) + \mathbb{E}_{c\in\mathscr{C}}\{f_c(\mathbf{u}_c, \mathbf{y}_c)\} \right\} \tag{2.1}$$

where $\mathbf{x} = \{\mathbf{x}_0, \mathbf{x}_c \text{ for } c = 1,\ldots,|\mathscr{C}|\}$ and similarly for $\mathbf{u}$,

subject to:

$$g_0(\mathbf{x}_0, \mathbf{u}_0, \mathbf{y}_0) = 0, \quad \text{pre-contingency power balance,} \tag{2.2}$$

$$h_0(\mathbf{x}_0, \mathbf{u}_0, \mathbf{y}_0) \le 0, \quad \text{pre-contingency operational limits,} \tag{2.3}$$

$$z_0(\bar{\mathbf{p}}, \mathbf{u}_0, \mathbf{y}_0) \le 0, \quad \text{feasible space for preventive controls,} \tag{2.4}$$

for $c = 1,\ldots,|\mathscr{C}|$:

$$g_c(\mathbf{x}_0, \mathbf{u}_0, \mathbf{x}_c, \mathbf{u}_c, \mathbf{y}_c) = 0, \quad \text{post-contingency power balance,} \tag{2.5}$$

$$h_c(\mathbf{x}_0, \mathbf{u}_0, \mathbf{x}_c, \mathbf{u}_c, \mathbf{y}_c) \le 0, \quad \text{post-contingency operational limits,} \tag{2.6}$$

$$z_c(\bar{\mathbf{p}}, \mathbf{u}_0, \mathbf{u}_c, \mathbf{y}_c) \le 0, \quad \text{feasible space for corrective controls.} \tag{2.7}$$

In (2.1–2.7), subscript $\mathbf{0}$ denotes the pre-contingency event (*i.e.*, the operation of the system before any outage happens) while subscripts $c \in \mathscr{C}$ are for different contingency events. Accordingly, the 1st term of objective function (2.1) denotes the costs of preventive controls that are firmly committed before the occurrence of any specific contingency. The 2nd term of (2.1) is the expectation of corrective control costs over a postulated set of credible contingency events. These latter costs will only be payed upon occurence of the specific contingency.

Parameter vector $\mathbf{y} = \{\mathbf{y}_0, \mathbf{y}_c \text{ for } c = 1,\ldots,|\mathscr{C}|\}$ represents the input bulk power system data that cannot be directly modified by the system operator in the timeframe of real-time operation (*e.g.*, the amount of power demanded by the system end-users and/or produced by renewable generation resources, the operational status and thermal rating of transmission branches, the capacity of dispatchable generation *etc.*). This vector varies between the pre-contingency event and different contingencies to denote a forced, *exogenous* change in the operational instance. We highlight that all parameters in vector $\mathbf{y}$ are in effect outputs of the cyber-layer of the electric power grid (most notably of the SCADA/EMS functionalities that gather remote measurement & sensing data to form the operator's perception of the grid operational instance).

Variable vector $\mathbf{x}$ denotes the system state (*i.e.* voltage magnitude and angle per node) while all control variables (*e.g.*, the amount of active power produced by generating units, the voltage setpoints of generating units, the position of switches and breakers of the transmission grid) are grouped in variable vector $\mathbf{u}$. Again, preventive control is with subscript $\mathbf{0}$ while corrective controls are with subscripts $c = 1,\ldots,C$. Notice the distinction between state and control variables here, as the values of the former are determined by the chosen values of the latter and the grid parameters as per the grid physical model. The physical model used for steady-state security management consists of the power flow equations (either in the non-linear AC format or the linearized DC power flow approximation) which are shown here as constraints (2.2) and (2.5) for the pre-contingency operation and all contingencies respectively.

Inequality constraints in (2.3, 2.6) express the engineering conditions that render the power grid physical operation acceptable, such as lower/upper voltage magnitude limits per node, power flow limits and voltage angle difference limits per branch, *etc.*. Inequality constraints (2.4,2.7) represent the feasible space of controllable resources (*e.g.*, the ramping rates and capacities of dispatchable generating units that are online), again for the base case and all credible contingencies respectively.

We use symbol $\bar{\mathbf{p}}$ to represent operation planning decisions that have been already implemented[1] and limit the currently admissible controls (*e.g.*, the commitment decisions of dispatchable generating units). Constraint group (2.7) also represents the physical restrictions linking preventive and corrective controls (*e.g.*, the ramping rates for generation redispatching, the maximum number of control actions that can be implemented by the operator in the considered time frame, *etc.*).

Constraints (2.2 – 2.7) express the security domain of an electric power system in real-time operation, with respect to the set of postulated contingencies $\mathscr{C}$. In other words, given such a set and a parameter vector $\mathbf{y}$, securing the system amounts to finding state and control variable vectors $(\mathbf{x}, \mathbf{u})$ satisfying (2.2 – 2.7). Ideally, this should be done at a minimum economic cost, expressed in objective function (2.1). For ease of notation, in the forthcoming parts of this report we will compactly write problem (2.1 – 2.7) as,

$$\begin{aligned} &\min_{\mathbf{x},\mathbf{u}} f_{\mathbf{u}}(\mathbf{u},\mathbf{y}), \\ &\quad \text{subject to:} \\ &\quad \mathbf{x} \in \mathscr{X}(\mathbf{u},\mathbf{y}), \\ &\quad \mathbf{u} \in \mathscr{U}(\bar{\mathbf{p}},\mathbf{y}). \end{aligned} \tag{2.8}$$

2.2.1 Operation in emergency mode

Decision making problem (2.1 – 2.7) formalizes the current practice of operating the system with a desired level of security, namely with the ability to withstand a pre-specified set of credible contingency events. While this level of security is mostly attainable during the day-to-day operation of the grid, under very rare and extreme circumstances this is not the case. For example, facing a severe weatherstorm that has physically destroyed several transmission lines in close proximity, the system may not be able to withstand the loss of an additional transmission component. Even though the practice of power system operators with regard to such situations is less documented and standardized, the overall philosophy is to use any means available (including, in the last resort, involuntary load shedding) to contain the extent of insecurity.

In mathematical programming terms, this translates into stating and solving a relaxation of problem (2.1 – 2.7) only in the event that it is found to be infeasible. For the purposes of this document we express such relaxation as:

$$\min_{\mathbf{x},\mathbf{u},\mathbf{r}} \Big\{ f_0(\mathbf{u}_0,\mathbf{y}_0) + f_{r0}(\mathbf{r}_0,\mathbf{y}_0) + \mathbb{E}_{c\in\mathscr{C}}\{f_c(\mathbf{u}_c,\mathbf{y}_c) + f_{rc}(\mathbf{r}_c,\mathbf{y}_c)\} \Big\} \tag{2.9}$$

subject to:

$$g_{r0}\left(\mathbf{x}_0,\mathbf{u}_0,\mathbf{r}_0,\mathbf{y}_0\right) = 0, \tag{2.10}$$

$$h_{r0}(\mathbf{x}_0,\mathbf{u}_0,\mathbf{r}_0,\mathbf{y}_0) \le 0, \tag{2.11}$$

$$z_0(\bar{\mathbf{p}},\mathbf{u}_0,\mathbf{y}_0) \le 0, \tag{2.12}$$

$$z_{r0}(\mathbf{r}_0,\mathbf{y}_0) \le 0, \tag{2.13}$$

for $c = 1,\ldots,|\mathscr{C}|$:

$$g_{rc}(\mathbf{x}_0,\mathbf{u}_0,\mathbf{r}_0,\mathbf{x}_c,\mathbf{u}_c,\mathbf{r}_c,\mathbf{y}_c) = 0, \tag{2.14}$$

$$h_{rc}(\mathbf{x}_0,\mathbf{u}_0,\mathbf{r}_0,\mathbf{x}_c,\mathbf{u}_c,\mathbf{r}_c,\mathbf{y}_c) \le 0, \tag{2.15}$$

$$z_c(\bar{\mathbf{p}},\mathbf{u}_0,\mathbf{u}_c,\mathbf{y}_c) \le 0, \tag{2.16}$$

$$z_{rc}(\mathbf{r}_0,\mathbf{r}_c,\mathbf{y}_c) \le 0, \tag{2.17}$$

where slack variable vectors $\mathbf{r} = \{\mathbf{r}_0, \mathbf{r}_c \text{ for } c = 1,\ldots,|\mathscr{C}|\}$ denote the relaxation of the system operational constraints in pre- and/or post-contingency mode. These variables correspond in practice to

[1] We use a bar on top of the letter $\bar{\mathbf{p}}$ to show that these values are fixed in the context of real-time operation.

applying involuntary load-shedding of the end-user demand and tolerating the violation of the operational limits of transmission components (*e.g.*, thermal ratings). The former measure is to be avoided to the extent possible, and if necessary, preferably only applied in the (rare) event of a contingency. It is obviously inevitable in problematic generation adequacy situations. Reasonably small violations of the operational limits of transmission assets are generally tolerable for a short period of time. However, load shedding may have to be used to relieve a large and prolonged violation that may otherwise trigger a cascading overload. Objective function (2.9) additionally includes non-negative functions f_{r0} and f_{rc} to account for the (extremely high) cost of operating the system outside its security domain.

Problem (2.9 – 2.17) is a generalization of (2.1 – 2.7) modeling the current practice of securing the system against a pre-defined contingency list, or doing a best effort to contain insecurity[2]. To simplify notation, in the forthcoming parts of this report we will compactly write problem (2.9 – 2.17) as,

$$\begin{aligned} &\min_{\mathbf{x},\mathbf{u},\mathbf{r}} f_{\mathbf{ur}}(\mathbf{u},\mathbf{r},\mathbf{y}), \\ &\quad \text{subject to:} \\ &\quad \mathbf{x} \in \mathscr{X}_r(\mathbf{u},\mathbf{r},\mathbf{y}), \\ &\quad \mathbf{u} \in \mathscr{U}(\bar{\mathbf{p}},\mathbf{y}), \\ &\quad \mathbf{r} \in \mathscr{R}(\mathbf{y}). \end{aligned} \tag{2.18}$$

2.3 Operation planning

Operation planning concerns taking decisions that must be enacted ahead of time, and have an effect over a horizon of several hours. Starting and stopping generating units is a prominent example as it can take several hours to warm-up a generating unit before it is ready to be connected with the grid and produce power while, once started, a unit may need to operate for several hours before it can safely cool down and be brought offline. This is the complementary security management problem to so-called real-time operation, solving ahead of time for the variables $\mathbf{p}$ that render the secure operation of the power grid possible at a reasonable socio-economic cost. Because of the proliferation of renewable power generation and the uncertainty it implies, the power grid operation planning problem is currently receiving considerable attention [10].

The classical formulation of the operation planning problem relies on a "best-guess" forecast for the values of exogenous random variables along the decision horizon ($\hat{\mathbf{y}}_t$ for $t = 1,\dots,T$) and only addresses the complexity associated with inter-temporal links between consecutive operational periods (*e.g.*, ramping rates for power generation, charging/discharging balances of energy storage *etc.*). Such a problem can be stated compactly as in (2.19 – 2.23).

$$\min_{\mathbf{p},\hat{\mathbf{x}},\hat{\mathbf{u}}} \Big\{ f_p(\mathbf{p}) + \sum_{t=1}^{T} f_{\mathbf{u}}(\hat{\mathbf{u}}_t,\hat{\mathbf{y}}_t) \Big\}, \tag{2.19}$$

subject to:

$$\mathbf{p} \in \mathscr{P}, \quad \text{planning decision feasible space,} \tag{2.20}$$

$$\hat{\mathbf{x}}_t \in \mathscr{X}(\hat{\mathbf{u}}_t,\hat{\mathbf{y}}_t), \text{for } t = 1,\dots,T, \tag{2.21}$$

$$\hat{\mathbf{u}}_1 \in \mathscr{U}(\mathbf{p},\bar{\mathbf{u}}_0,\hat{\mathbf{y}}_1), \tag{2.22}$$

$$\hat{\mathbf{u}}_t \in \mathscr{U}(\mathbf{p},\hat{\mathbf{u}}_{t-1},\hat{\mathbf{y}}_t), \text{for } t = 2,\dots,T. \tag{2.23}$$

[2] Even though (2.9 – 2.17) was introduced in the context of a system operator taking decisions from the control room in a centralized manner, notice that this problem statement is (intentionally) abstract enough to also allow alternative interpretations, concerning the automata of the power grid that perform decentralized control policies on a local scale. In the remainder of this report we formalize cyber-physical power grid security management as a stochastic decision making problem on the basis of (2.9 – 2.17) and while verbally referring to the centralized functionalities of the system operator. In principle, the extension to the functionalities performed by distinct control & protection automata of the power grid is straightforward.

Objective function (2.19) jointly minimizes the cost of operation planning decisions along with the cost of real-time operation along the best-guess forecast. Constraint (2.20) expresses the set of possible operation planning decisions (*e.g.*, minimum up-/down- times of generating units). Further, constraints (2.21 – 2.23) enforce that operational planning decisions should render the real-time operation security management problem feasible. Note that in (2.22 – 2.23) we have augmented the symbolic expression for the admissible set of real-time controls by explicitly showing its dependence on past decisions[3]

The acknowledgment of the uncertainty concerning exogenous random processes in $(\mathbf{y})$ implies an explosion of computational complexity. The quantities of interest are continuously-valued (*e.g.*, wind-power generation) implying infinite-dimensional optimization problems while the gradual resolution of the uncertainty (*e.g.*, the realized wind forecasting errors at 9 AM do not uniquely determine the errors for 9 PM) call for non-anticipativity constraints between alternative realizations. To date, the most pragmatic approach for handling such problems invokes the two-stage, sample average approximation. It amounts to sampling a finite set of trajectories $\mathbf{y}^{\xi}$, for $\xi = 1,\ldots,\Xi$, and assuming that the uncertainty characterizing the complete horizon is resolved at beginning of the horizon. The former reduces the infinite-dimensional problem to a (tractable) finite-dimensional one while the latter implies enforcing coupling constraints only along a given trajectory $\xi = 1,\ldots,\Xi$. Given a set of trajectories, problem (2.19 – 2.23) can be extended as:

$$\min_{\mathbf{p},\hat{\mathbf{x}},\hat{\mathbf{u}}}\Big\{f_p(\mathbf{p}) + \sum_{\xi=1}^{\Xi}\lambda^{\xi}\sum_{t=1}^{T} f_{\mathbf{u}}(\mathbf{u}_t^{\xi},\mathbf{y}_t^{\xi})\Big\}, \tag{2.24}$$

subject to:

$$\mathbf{p} \in \mathscr{P}, \tag{2.25}$$

$$\mathbf{x}_t^{\xi} \in \mathscr{X}(\mathbf{u}_t^{\xi},\mathbf{y}_t^{\xi}), \text{for } t = 1,\ldots,T \text{ and for } \xi = 1,\ldots,\Xi, \tag{2.26}$$

$$\mathbf{u}_1^{\xi} \in \mathscr{U}(\mathbf{p},\bar{\mathbf{u}}_0,\mathbf{y}_1^{\xi}), \text{for } \xi = 1,\ldots,\Xi, \tag{2.27}$$

$$\mathbf{u}_t^{\xi} \in \mathscr{U}(\mathbf{p},\mathbf{u}_{t-1}^{\xi},\mathbf{y}_t^{\xi}), \text{for } t = 2,\ldots,T \text{ and for } \xi = 1,\ldots,\Xi. \tag{2.28}$$

where parameter λ^{ξ} is a weighting factor for the relative importance of the different trajectories (*e.g.*, $\lambda^{\xi} = 1/\Xi$ for uniform weighting). In comparison to problem (2.19 – 2.23), problem (2.24 – 2.28) repeats all constraints and variables related to the real-time operation stage over all postulated trajectories. Thus, the computational burden of such a problem grows quickly with the number of postulated trajectories Ξ.

We must finally acknowledge that, as evidenced in [10], there is a strong emerging research stream concerning chance-constrained power grid operation planning. These approaches seek to ensure that the system operation is achievable at least with a given probability (in other words, a probabilistic guarantee that planning decisions render the system operable). The mathematical foundation includes assuming a probability distribution and analytically reformulating the problem constraints according to its properties. While chance-constrained optimization is a very interesting and relevant direction for operation planning security management, there are several issues that must be underlined here. At present, the most advanced approaches in the literature rely on much simplified modeling of the power grid operation, thus further research is needed. In light of this fact, and taking the scope of the CYPRESS project into account, we selected here not to emphasize such approaches.

[3] The feasible space for real-time control actions to be applied at the first period of the decision horizon ($t = 1$) depends on the real-time control actions applied before the start of this horizon. We represent such dependence in (2.22) using $\bar{\mathbf{u}}_0$ to denote the anticipate real-time control actions that would be applied up to the start of the considered decision horizon.

3

Cyber-physical threats and countermeasures

This chapter concerns the cyber-physical threats facing the electric power grid and countermeasures that can be applied on the cyber sub-system to address these. Referring the reader to the 2nd report of the CYPRESS project [11] for a more detailed description of such threats and countermeasures, we focus here on expressing these concepts in a generic mathematical model.

3.1 Taxonomy of cyber-physical threats

Cyber-physical threats are malfunctions of the cyber sub-system with the potential to adversely affect the operation of the electric power grid. The broad range of functionalities provided by the cyber sub-system for the operation of the power grid implies a broad range of such threats. Recent survey works [2]–[6] analytically present this broad range of threat and countermeasure instances and introduce alternative classification schemes.

Comprehensively describing a cyber-physical threat amounts to defining the role and motivation of a potential threat agent (*i.e.*, the actor potentially posing the considered threat to the power grid), the resources of the threat agent and, last but not the least, the threat mode (*i.e.*, the way in which the functionality of the cyber sub-system is compromised to put the power grid at risk). Adversary roles are already well covered in the previous reports of the CYPRESS project (see section 4.3.4 in [7] and/or 3.1.3 in [11]). Here, it is important to acknowledge that both malicious adversaries as well as non-malicious actors (*e.g.*, erroneously trained employees, software or hardware vendors) may assume the role of a threat agent. It is also important to acknowledge that the electric power grid is also at risk from cyber-physical threats with no (obvious) associated threat agent. Indeed, the natural (un)reliability of the cyber sub-system components is a threat to the electric power grid security

without an associated agent.

Defining the resources of a threat agent is arguably a non-trivial exercise. A commonly used approach is to focus on the worst-case threat posed by a computationally powerful and knowledgeable adversary that has performed diligent reconnaissance. It may however be relevant to also consider alternative cyber-physical threats, differentiated by the adversary's computational resources and the quality of information he has on the attacked system. In fact, game-theoretic analysis of the defender vs cyber-attacker interaction suggests that preparing to face a rational adversary with unbounded computational resources and with perfect information does not necessarily protect from alternative attack strategies launched by (realistic) adversaries with bounded computational resources and/or imperfect information [12]. Further, research works have shown that an adversary does not need complete or fully correct knowledge of the power grid to launch a successful cyber-physical attack [13], [14].

Concerning the mode of a cyber-physical threat, we adopt a scheme inspired from [5] and will henceforth distinguish between *control-signal based, measurement based* and *coordinated (measurement & control-signal based)* threats.

Control-signal based threats compromise the integrity and/or availability of information flow *towards* field devices, so as to directly execute control actions that jeopardize the physical operation of the power grid. The most famous *AURORA (i.e., Avoiding Unwanted Reclosing of Rotating Apparatus)* is a prominent example. It involves a series of malicious switching commands on the circuit breaker protecting a generating unit, to produce great torque and electrical stress and physically destroy the rotating machine [15]. We note that this class of threats also includes threats compromising the availability of information, *e.g. Denial-of-Service (DoS)* attacks that deprive the system operator of the ability to remotely control the substation switchgear.

Measurement based threats compromise the integrity and/or availability of information collected *from* the field devices, so as to induce erroneous operation of automated control loops and/or deceive the system operators into performing erroneous actions or missing alerts. The so-called *AGC attack* [16] is a typical example of inducing erroneous operation of an automated control loop. The threat concerns tampering with the input measurements for the *Area Control Error (ACE)* computation, to induce maloperation of the grid frequency maintaining mechanism (*i.e.*, *Automatic Generation Control*). It can bring the grid to instability, which would imply severe consequences to its end-users. The well-studied *load redistribution attack* (see, for instance, [13], [14], [17]–[21]) is another example. It concerns presenting false load data to the system operator so as to provoke misguided reactions, in fine causing economic efficiency losses and/or physical security violations. This class of threats also includes threats compromising the availability rather than integrity of information. An example is the incident of March 2019, wherein as reported by the *North American Energy Reliability Corporation (NERC)* [22] malicious cyber-attackers managed to disrupt the communication between remote field devices and a control centre, thus depriving the system operator of her situation awareness.

Coordinated (measurement & control-signal based) threats are sophisticated malicious attacks that combine multiple threat modes belonging in the aforementioned threat classes. For example, cyber-attacks including a change in the power grid topology through taking control of breakers & switches (control-signal based threat) and the injection of false-data to hide this event from the system operator (measurement based threat) have received considerable attention in the literature (indicatively, see [20], [23]–[25]). Further, the analysis of the known power grid cyber-physical insecurity incidents in Ukraine confirms that some real-life cyber-physical attacks have indeed coordinated threats manipulating both measurements as well as control functionalities [26], [27].

3.1.1 Selected examples from the literature

As a central component of the power system, the supervisory control and data acquisition (SCADA) system can be regarded one of the primary targets for cyber-attacks. Feeder automation (FA) is one of the crucial applications of the SCADA system. FA system can locate the faults, and realize isolation and power restoration with the aid of the cyber and physical components. The remote terminal units (RTUs) gather the data from the switches and other physical equipment and send them to the master station through the communication network. The master station enables the operators to monitor and control the system based on the advanced software. The RTUs are often installed in a remote area and lack effective protection.

In [28], the impact of the cyber-attack on the operation of FA system is investigated, by exploiting the vulnerabilities in RTUs. After obtaining the control privilege on RTUs, the attacker injects false information to make an undesired FA operation. Another cyber-attack on the SCADA system has been investigated in [29]. In this work, a semi-Markov process is employed to model and evaluate the cyber-attacks against the SCADA systems considering insider assistance. Also, a FlipIt game model is presented to assess the defense and attack strategies, and analyze the impacts of insider assistance.

Beyond the central function of the SCADA system, substations are also potential targets for malicious cyber-physical attacks. Once the attacker gains access to a substation by compromising the network firewall, he/she can inject abnormal control signals or manipulate sensor data measurements that can result in unnecessary actuation of the switching devices. For example, a slight modification of the settings of a relay in a protection assembly can open the slave circuit breaker and isolate the linking transmission line from the power system [30]. In the formal dynamic attack, the attacker maximizes the system damage via the budget-constrained attacks at different time constants to strategically identified critical substations.

Another significant operational function in power systems that can be regarded as a potential target for cyber attacks is *Load Frequency Control* (LFC). The LFC signals are distributed to the generators, so as to ensure power balance and guarantee frequency stability. By manipulating the variables in LFC, attackers can destabilize the balance of active power and cause cascading failures. In [31], only cyber attacks on tie-line interchange power measurement are considered, and only one area is assumed to be compromised. In addition, a level of imperfectness is considered in both attack and defense sides. In this study, the attacker has two attacking options of false data injection (FDI) strategies:

- Damage oriented FDI attack: In this case, the attacker injects false data to increase the attack physical impact, which however increases the possibility of being detected by the detector.
- Deception oriented FDI attack: In this case, the attacker tries to make the attacked value of the monitoring variable below the detection alarm value, in order to avoid being detected by the alarm.

3.2 Taxonomy of countermeasures acting on the cyber sub-system

The development of countermeasures to safeguard the functionality of the cyber sub-system of the electric power grid is continuously evolving [2]–[6]. For the sake of formalizing a generally applicable decision-making approach, we will categorize such countermeasures as serving the functions of *protection, detection* and *mitigation*.

Protection countermeasures can be used at the level of individual field devices, parts of the cyber sub-system as well as the integrated cyber-physical system. At the level of field devices, consider the example of *security and encryption techniques* used to make the system immune to measurement based attacks. *Firewalls* blocking potentially harmful traffic from penetrating the communications network are a well-established tool at the level of a part of the cyber sub-system. Also, several redundant and independent data acquisition and state estimation layers could be used to protect the

SCADA/EMS functionalities. Finally, an interesting protection countermeasure at the level of the integrated cyber-physical system consists of preserving the confidentiality of the cyber and/or physical infrastructure properties (e.g., the topological information of the transmission grid). Indeed, depriving potential adversaries from the information required to design a successful cyber-physical attack strategy effectively protects the integrated system from such a threat. The concept of *moving target defense* involves frequent, relatively minor adaptations of the system configuration so as to compromise the knowledge of a potential adversary [32].

Detection countermeasures are a necessary complement to protection, since it is practically impossible to guarantee that no cyber-physical threat will ever materialize. The so-called *Intrusion Detection Schemes* (IDSs) operate at the level of the communication network and are commonly either *signature* based or *anomaly* based. Signature based schemes rely on physical watermarking, *i.e.* injecting a known noise on top of an input signal so as to produce a predictable output, for verifying the integrity of communications. Anomaly based schemes monitor the network traffic and rely on filtering techniques (and, more recently advanced Artificial Intelligence/Machine learning tools) to identify and flag suspicious traffic patterns, suspected bad data and suspected false data[1]. Further, at the level of the integrated cyber-physical system model-based detection can also be applied. Model-based schemes rely on simulating the estimated evolution of the physical system (e.g., near real-time, look-ahead power flow analysis) so as to identify suspicious deviations in the information and data produced by the cyber sub-system (*e.g.*, bad data identification techniques on top of model-based state estimation). It must of course be noted that detection measures only serve to identify the occurrence of a cyber-physical threat rather than stop it from compromising the functioning of the electric power grid.

Mitigation countermeasures serve to neutralize the effect of a specific threat whose occurrence has already been detected (using the tools introduced in the previous paragraph) on the functioning of the cyber sub-system (and, in consequence, on the electric power grid). The most commonly documented techniques to date only concern a part of the cyber sub-system, namely the communication network. For instance, pushback methods can be deployed to block incoming traffic from possibly compromised nodes of the communication network and reconfiguration methods can be used to remove possibly compromised network nodes.

3.2.1 Selected examples from the literature

The section introduces several references that represent the forefront of studies in the tri-level modeling of defender-attacker-defender games for cyber attacks across various domains of power systems. These references were chosen for their significance in illustrating how to model attack budget, uncertainty, and other critical factors. It is crucial to emphasize that these references encompass various aspects, including uncertainty on both the attacker and defender sides, as well as innovative techniques for modeling and solving optimization problems.

In [30], a game-theoretic approach is presented to design a new attacker-defender model in power systems for dynamic cyber-attacks on protection systems. Given a power system network, the defender attempts to minimize the load loss in response to the inserted dynamic attacks at different time instants. Considering the defense budget, the defender strategically identifies the set of critical substations for protection.

A three-stage game-theoretic framework for cyber-attacks on SCADA systems is proposed in [28]. In the defense stage, the defender tries to protect the RTUs with extra security resources regarding the defense budget constraints. In addition, in the recovery stage, the master station tries to

[1] Whereas bad data may be the result of an unwanted error in a measurement or computation process, false data are intentionally introduced by a malicious adversary and typically designed with plausible realistic properties so as to deceive known bad data identification techniques.

isolate the faulty area based on the false fault information and restore the power supply as much as possible.

The defender invests security resources in the RTUs to minimize expected economic impacts. In the recovery stage, when the master station receives the fault information generated in the attacking stage, the FA system operates to minimize the impacts of the attack. The recovery stage consists of three steps that are optimally determined by the FA system. Regarding uncertainty and misinformation, the FA is applied with the Bayesian inference-based fault-tolerant location technique. The corrective actions of the FA system are (i) detecting fault location, (ii) fault isolation and (iii) power supply restoration.

In [33], a new method is proposed for the optimal allocation of limited defensive resources to safeguard the power grid against load redistribution attacks. In this method, the security administrators install intrusion detection systems (IDS) in the substations, to both detect and block intrusions. Due to the defense budget limitations, only a few critical substations can be chosen for IDS installation. To effectively select the critical substations, both the topological information of the power grid network and the operation conditions of the power system shall be considered. The following criteria should be satisfied in the determination of critical substations:

1. Should the authenticity of the measurements in a critical substation be guaranteed, no load loss could be inserted by the LR attacks, even in the highest loading situation of power systems.

2. LR attacks can be withstood, even upon the successful defeat of their IDS, due to the existence of redundancy in the critical substations.

In light of the reviewed literature, power system operators commonly allocate limited resources for the protection of critical substations identified during the defensive stage. This observation is drawn from the consistent consideration of a defense budget in the most reviewed papers. Therefore, as a common practice for power system operators, limited resources for the protection of critical substations selected in the defensive stage are distributed to upgrade the security measures in the substations (such as IDS and firewalls). Such resource allocation leads to an improvement in the security level of the updated substations, and, a lower probability of being successfully compromised, as a result. The power grid security administrators (defenders), at the lower level, determine the optimal remedial strategies based on OPF to minimize the consequences of the attack.

In [34], a zero-sum game between the attacker and defender is modeled for indirect cyber-attacks on the electricity market. This game defines the proportion of times that the attacker and defender attack and defend different measurements, respectively. In the designed model, the objective of the defender is to prevent measurements from being changed by the attacker. The variables on the defensive side are the indices of protected measurements, and constraints define the defense budget. In the designed multistage optimization framework, the defender tries to decrease the change in measurements by protecting some of them that have higher risk values.

In [35], a new false pricing attack and model for interaction between attackers and defenders is presented, using a zero-sum Markov game. This false pricing attack can be considered a direct cyber-attack on the electricity market and is modeled as a tri-level optimization problem. In the planning stage of this tri-level optimization problem, the defender tries to harden and protect critical smart meters against false price injection attack, regarding the defensive budget. In this stage, the objective of optimization is to find critical meters to be hardened. Also, in the operating stage, in case of overloading lines and generators, the operator responds immediately to the new state of the system and sheds some loads to minimize the consequences of the attack. In this stage, the optimization variables are indices of load points under load shedding.

In this optimization problem, the objective function is the total interrupted load in the system, that the attacker and defender try to increase and decrease, respectively. The defensive constraints are the defense budget, limitations for the amount of load shedding at each bus, the amount of transmitted power through lines, the demand response rule based on falsified prices, and the resistance of protected meters against manipulation.

In [31], a game theory model is introduced to analyze the optimal strategy of attack-defense interaction for load-frequency control systems. In the designed interaction model, the defender has two kinds of detection schemes:

- Alarm-based detection: In this case, the defender uses false data alarm to distinguish between attacked and safe signals. If the value of ACE signal exceeds the alarm value set, then the attack is detected; otherwise, the attack will be undetectable.
- Threshold-based detection: In this case, detection works based on the anomaly of transient behaviors of attacked variables.

It is crucial for the defender to choose the appropriate alarm and threshold values to cope with possible attack strategies. In this study, the pay-off defense function is selected as the objective function of the problem, in which the attacker tries to minimize its value and the defender tries to maximize it. The pay-off defense value is the difference between defense and attack utility functions. The attack utility function is proportionate to the precision function and the defense utility function is proportionate to the cumulative frequency deviation.

A cyber-physical coordinated defense strategy is designed in [36], to overcome the disruption and minimize the risk of coordinated cyber-attacks on Smart Grids. In this research, the defender protects transmission lines at the cyber layer by allocating defensive resources that can reduce the success probability of topological attacks. If the attacks succeed, then the defender performs the modified security-constrained optimal power flow, in the recovery stage. The defender must perform a re-dispatch to avoid major blackouts, including regulating the power output of generators and shedding parts of loads.

At the cyber layer, the defender distributes defensive resources to cyber components. In the case of protecting a single component, its associated lines are also protected. Therefore, the successfully attacked probabilities of these lines are lowered. Between two successive attacks, the defender tries to detect and analyze anomalies, recognize the performed attacks and allocate extra resources to protect other vulnerable lines. For instance, if the attacker has tripped the first line, the defender may detect its outage and predict that the second and third lines are vulnerable, so he/she allocates extra defensive resources to these lines.

After primary outages due to cyber-attack occurrence, the power flow of the tripped lines will be redistributed, so other lines may be overloaded. To eliminate cascading outages, the defender must perform SC-OPF, which may need load shedding in some load points. If the attacks succeed, the defender tries to ramp up/down generators and shed uninterruptible load to reduce overloading. If the overloading condition still remains, the defender has to shed uninterruptible loads to maintain system reliability and stability.

4

Stochastic programming for power system cyber-physical security management

This chapter concerns the extension of the multi-stage stochastic programming approach for physical security management introduced in chapter 2, so as to also take into account the cyber-physical threats and countermeasures discussed in chapter 3. This extension is particularly challenging due to: (i) the need to additionaly model the cyber sub-system functionalities on top of the power grid physics and (ii) the fact that cyber-physical threats are often methodically planned by independent adversarial decision makers.

In order to tackle the first challenge, section 4.1 introduces a modeling abstraction of the power system cyber-layer as the interface between a decision maker (be it the power system planner/operator or an adversary) and the physical processses of electricity generation, transmission and distribution. The abstraction allows to represent all threat & countermeasure classes from chapter 3 in the compact statement of the security management stochastic program. The remainder of this chapter focuses on the second challenge. Section 4.2 frames the power system cyber-physical security management problem in question, by introducing the considered timeframes, the considered decision making actors and their interaction with the power grid cyber and physical layers. Next, section 4.3 introduces a model for a cyber-physical attacker as an independent decision making agent and finally section 4.4 develops alternative formulations for the cyber-physical security management problem from the planning perspective.

4.1 Cyber sub-system modeling abstraction

We consider the cyber sub-system of the electric power grid as an interface between its physical processes and the power system operator. Indeed, centralized functionalities of the cyber sub-system (notably, the SCADA/EMS) are in place to enable the operator to oversee and control the physical processes of electricity generation, transmission and distribution. Decentralized and distributed control and protection functionalities are in place to locally implement the control policies pre-defined by the power system operator in an automated or semi-automated manner. Finally, measurement/sensing instrumentation as well as communications infrastructure is in place to service this interface.

4.1.1 Cyber sub-system interface variables

In order to express the cyber sub-system modeling abstraction, we rely on two complementary interface variables, holding operation-level and field-level values describing the power grid physical processes. The values of the operation-level interface variables $\mathbf{y}$ are the values describing the physical processes that are perceived by the power-grid operator through the SCADA/EMS functionalities. The values of the field-level interface variables $\mathbf{w}$ are the values perceived by the field devices of the power grid. The power-grid cyber sub-system then interfaces the physical processes and the power system operator by:

- Measuring/sensing the physical processes, collecting, communicating and processing data to produce the values of operation-level variables $\mathbf{y}$ that the power grid operator will use as input to supervise and control the system.
- Enacting the protection and control functions decided by the power grid operator to produce desired values of the field-level variables $\mathbf{w}$ for the electric power grid physical processes.

A well functioning cyber sub-system should ensure both that operational-level values truthfully represent the physical processes and that field-level values comply with the power grid operator's protection and control decisions. Let us consider by way of example any transmission branch of the electric power grid. It may be out of service, *i.e.* field-level value equal to zero, due to the natural unreliability of its components. The power grid operator relies on the cyber sub-system to accurately sense the branch breaker positions and in fine translate these into a zero operation-level value for the branch status. In an alternative case, the operator may decide to modify the topology of the grid by removing a branch from service. To do so the operator can only change the operation-level branch status value to zero and rely on the communication and control functionalities of the cyber sub-system to change the branch breaker positions so that the field-level branch status value is also set to zero.

Let us now briefly return to the statement of the physical security management stochastic programming problems from Chapter 2 and the so-called *situation awareness* parameter vector $\mathbf{y}$. As mentioned, it holds the input data for the decision-making problem of the power grid operator and is an output of the power-grid cyber sub-system. In the context of *cyber*-physical security management this vector $\mathbf{y}$ is in effect the operation-level interface variable. The field-level interface variable does not appear in any of the statements included in Chapter 2. Indeed, the physical-only approach for power system security presupposes that the cyber sub-system is perfectly reliable, hence the operation-level variable is sufficient to represent information from the physical processes. We will henceforth use symbol $\mathbf{w}$ for the field-level interface variable.

4.1.2 Cyber sub-system threats & countermeasures

We start by postulating that the cyber sub-system of the electric power grid can be characterized by an (internal) cyber sub-system state variable vector, in a similar manner as the electrical state variable vector (voltage magnitude and angle per node) characterizes the physical power grid. Cyber-physical

threats can be understood as alterations of the cyber sub-system state, which are liable to jeopardize the power grid security. Rather than explicitly modeling such cyber-state alterations, we choose to directly represent cyber-physical threats as modifications of the variables $\mathbf{y}$ and/or $\mathbf{w}$ at the interface of the cyber sub-system and the electric power grid. More specifically, we reserve subscript $\emptyset$ to exclusively denote the event wherein the cyber sub-system of the electric power grid is operating as intended. For any alternative threat vector $\mathbf{a} \in \mathscr{A}$ altering the state of the cyber sub-system we write:

$$\mathbf{y}(\mathbf{a}) = \mathbf{y}_{\emptyset} + \mathbf{dy}(\mathbf{a}), \tag{4.1}$$

$$\mathbf{w}(\mathbf{a}) = \mathbf{w}_{\emptyset} + \mathbf{dw}(\mathbf{a}), \tag{4.2}$$

while denoting by $\mathbf{dy}(\mathbf{a}), \mathbf{dw}(\mathbf{a})$ the effect of the threat a on the values of operation-level and field-level interface variables.

Returning to the previous example, let us consider that transmission branch i is originally in service. The i-th element of the interface variable sub-vectors referring to the status of the grid transmission branches would by default be $\mathbf{y}_i = \mathbf{w}_i = \mathbf{y}_{\emptyset,i} = \mathbf{w}_{\emptyset,i} = 1$. Consider now that threat $\mathbf{a}$ is a measurement based attack corrupting communications so as to convey false breaker positions towards the control centre. It could be represented by way of a change in the operation-level interface variable only as in,

$$\mathbf{w}_i \leftarrow \mathbf{w}_i(\mathbf{a}) = \mathbf{w}_{\emptyset,i} = 1, \tag{4.3}$$

$$\mathbf{y}_i \leftarrow \mathbf{y}_i(\mathbf{a}) = \mathbf{w}_i + \mathbf{dy}_i(\mathbf{a}) = 0. \tag{4.4}$$

Indeed, in (4.3) the field-level variable describing the branch status does not change since the threat has only altered the measurement value that is shared towards the power grid control centre, not the position of the branch breakers.

Returing to the default $\mathbf{y}_{\emptyset,i} = \mathbf{w}_{\emptyset,i} = 1$, consider next a control based attack $\mathbf{a}'$ exploiting remote unauthorised access to a substation automation system so as to open the branch breaker. Equality (4.5) could be used to represent the actual change of the breaker position (field-level variable) and, assuming that communications are not highjacked, equality (4.6) would accordingly update the operation-level interface variable.

$$\mathbf{w}_i \leftarrow \mathbf{w}_i(\mathbf{a}') = \mathbf{w}_{\emptyset,i} + \mathbf{dw}_i(\mathbf{a}') = 0, \tag{4.5}$$

$$\mathbf{y}_i \leftarrow \mathbf{y}_i(\mathbf{a}') = \mathbf{w}_i = 0. \tag{4.6}$$

Finally, a coordinated (measurement & control based) attack $\mathbf{a}''$ can conceivably both remove the branch from service (by exploiting access on the substation automation system) and conceal this change from the system operator (by compromising communications). In our modeling abstraction, it is representable as follows.

$$\mathbf{w}_i \leftarrow \mathbf{w}_i(\mathbf{a}'') = \mathbf{w}_{\emptyset,i} + \mathbf{dw}_i(\mathbf{a}'') = 0, \tag{4.7}$$

$$\mathbf{y}_i \leftarrow \mathbf{y}_i(\mathbf{a}'') = \mathbf{w}_i + \mathbf{dy}_i(\mathbf{a}'') = 1. \tag{4.8}$$

As mentioned in Section 3.2 a variety of hardware and software tools is already available to protect the cyber sub-system of the power grid from intrusion & manipulation, to detect unwanted events and to mitigate their impact on its functionality. We seek to model the effect of any such measure at the interface between the cyber sub-system and physical power grid. Variable vectors $(\mathbf{y}, \mathbf{w})$ represent this interface. Let us now denote as $\mathbf{m}$ the decision corresponding to the properties of the countermeasures acting on the cyber sub-system (*e.g.*, setting the rules of existing firewalls, the thresholds for bad-data detectors, the topology of the communications network *etc.*). In the general case, such decision determines the subset of threats against which the cyber sub-system is secured as well as the resulting degree of confidentiality, integrity and availability. To represent this, we edit our notation as follows:

$$\mathbf{dy}(\mathbf{a}) \leftarrow \mathbf{dy}(\mathbf{m}, \mathbf{a}) \tag{4.9}$$

$$\mathbf{dw}(\mathbf{a}) \leftarrow \mathbf{dw}(\mathbf{m}, \mathbf{a}). \tag{4.10}$$

For a concrete example of (4.9) consider the use of redundant data acquisition layers concerning the nodal power injection quantities. Using this measure, the components of vector $\mathbf{y}$ that hold the nodal power injection measurements cannot be modified by means of attacks on any single layer of the data acquisition infrastructure. Similarly, for a concrete example of (4.10) consider the use of a firewall blocking all incoming traffic towards a specific substation of the power grid. Using this measure, the components of vector $\mathbf{w}$ that hold the substation switchgear positions cannot be modified by threats acting on the cyber sub-system that involve remote connection.

Designing a perfectly effective countermeasure scheme to protect the cyber sub-system of the power grid against any conceivable threat is obviously not realistic. In order to represent technical limitations, we will henceforth restrict the choice of countermeasures within a corresponding set $\mathbf{m} \in \mathscr{M}$. Finally, we will use function $f_m(\mathbf{m})$ to account of the cost associated to the countermeasures acting on the cyber sub-system of the electric power grid.

4.2 Problem description

We place ourselves in the operation planning context and seek to identify the (*here-and-now*) decisions that have to be taken ahead of the real-time operation of the electric power grid. In order to keep the exposition of the *cyber-physical* security management problem as simple as possible, we consider planning with respect to a single operational period and given a single "best-guess" forecast for the values of exogenous random variables concerning the physical system, such as renewable power generation and load demand. In other words, planning uncertaintly only relates to the potential development of a malicious cyber-physical attack against the electric power grid. At the same time, real-time operation remains in itself a decision-making problem under uncertainty regarding the potential occurence of contingencies.

Figure 4.1 graphically represents the considered decision-making instance. The yellow light-bulb graphically represents a functioning physical layer of the electric power grid whereas its cyber sub-system is represented as the black computer terminal. As shown in figure 4.1, an instance of cyber-physical security management decision-making consists of sequential decisions taken by three independent actors. In inverse chronological order, the latest decisions belong to the real-time operator (in green). She relies on the cyber sub-system to interface the physical layer of the power grid and ultimately secure its functionality. The decisions of an *attacker* (in red) are assumed to precede those of the real-time operator. The attacker has to decide how to interact with the cyber sub-system of the electric power grid so as to disrupt the electricity supply service. Earlierst in time, we consider an integrated planner (in blue) that may act both on the physical layer of the electric power grid as well as its cyber sub-system.

The reader may well understand that the decision-making problem of any one of the aforementioned actors is a decision-making problem under uncertainty that can be framed in the language of multi-stage stochastic programming. Indeed, the real-time operator faces uncertainty on the potential occurence of contingency events as well as on the effectivenes of post-contingency corrective controls [37]. Acknwoledging that the cyber sub-system is vulnerable to cyber threats (and notably malicious attacks), she may eventualy also adapt her decision-making approach to explicitly account for the uncertainty on the state of the cyber sub-system. An attacker on the other hand, can only decide based on the models and data she has at her disposal. She would therefore face the uncertainty regarding the accuracy of her models and assumptions. Moreover, as she has to decide ahead of real-time, she also faces uncertainty on the exogenous factors driving the power system operation (*e.g.*, the wind/solar power generation). All the aforementioned uncertainties, along with the random properties of independent malicious adversaries, characterize the decision-making problem of the planner.

We choose to focus hereafter on formulating a decision-making problem under uncertainty from the perspective of the integrated cyber-physical planner. Our choice is motivated by the fact that this is the most comprehensive decision-making problem in question since requiring to anticipate the

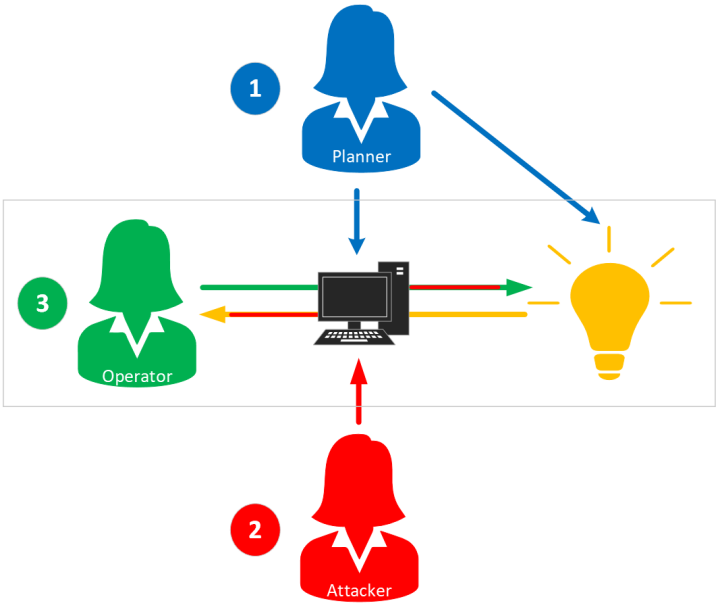

Figure 4.1: Sequence of decisions of cyber-physical security management

future decisions of both the cyber-attacker and real-time operator. Notice that this choice is well-aligned with the change management practice in electric power systems, wherein new approaches are typically easier to put in practice in the operation planning context with respect to the context of real-time operation. It is also motivated by our current understanding of the decision timeframe for cybersecurity countermeasures. Figure 4.2 illustrates the main features of the decision making models of the independent agents involved in cyber-physical power system security management.

At the latest decision making opportunity of real-time operation, we consider that the system operator would make a best effort to maintain the grid within its physical security limits (and while assuming that the cyber sub-system of the power grid is fully functional). We represent this in Figure 4.2 by only including the light-bulb symbol (for the power grid) in the thinking bubble of the power grid operator. Notice the green outline of this light-bulb symbol. It is meant to represent that the power grid operator will make her decisions according to her own model of the physical power grid. This latter model may well be incorrect/inaccurate, for instance in case the cyber-physical adversary succeedes in launching a cyber-physical attack. In any case, we use the general statement (2.18) as the decision-making model for the real-time operator of the power grid.

The decision-making model of the cyber-attacker includes both the physical power grid (that she may target to disturb) and its cyber sub-system (that she can use as a medium). Notice that both symbols for the power grid (light-bulb) and cyber sub-system (black desktop computer) are shown with a red outline in the thinking bubble of the cyber-physical adversary. We do so to represent that the adversary can only rely on her own models for the physical grid and its cyber sub-system, which may be incorrect/inaccurate either because the adversary has not performed diligent reconnaissance and/or because of deception/misinformation countermeasures employed by the cyber-physical security manager. Additionaly, the adversary needs to anticipate the future decisions of the real-time operator, as these in turn will determine the actual effect of a threat vector on the electric power grid. We therefore also include a model for the power grid operator inside the adversary's thinking bubble, with the red outline once again implying that this model may be incorrect/inaccurate. In any case, we assume that each individual cyber-attacker is a deterministic decision-maker and develop the generic statement of the corresponding problem in the following section 4.3.

The agent acting first and shown in blue is the integrated cyber-physical planner. The decision making model of this agent also includes both the physical power grid (that she is targeting to secure) and its cyber sub-system (that she may need to protect). We use a blue outline on the respective

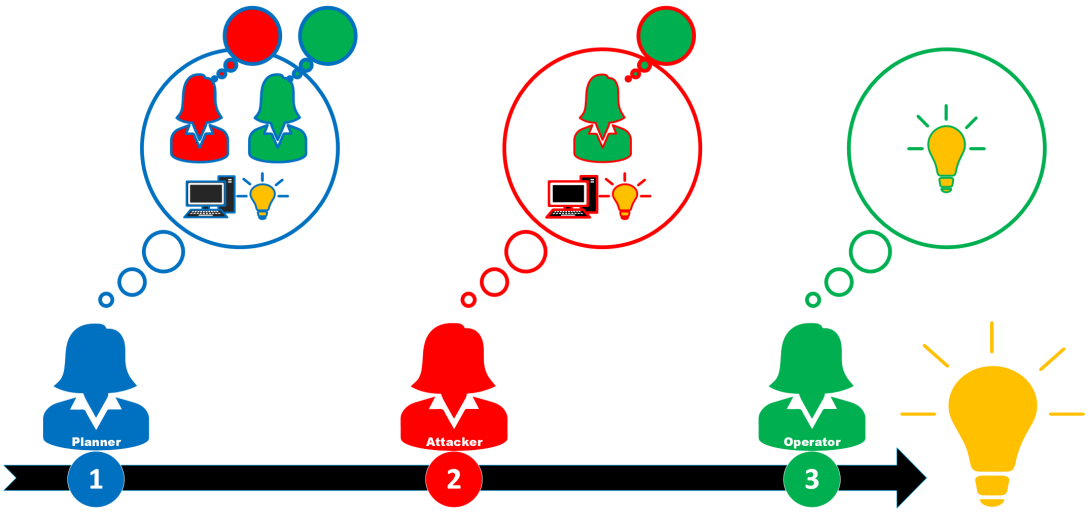

Figure 4.2: Overview of decision-making problems

symbol, to represent the models that are of interest to the planner, possibly of different accuracy and/or level of detail with respect to the models of the cyber-attacker and of the power grid operator. This agent is required to anticipate the distinct actions of (i) alternative cyber-physical adversaries that may launch a threat vector, and (ii) the power grid operator that will take the last decisions on how to operate the grid. We therefore embed two decision making sub-layers in the model of the cyber-physical security manager. The first one, in red, corresponds to the cyber-physical attacker(s) that may observe (partially or fully) the decisions of the security manager and decide to launch a threat vector while anticipating what the power grid operator would do. The second one, in green, represents the power grid operator as anticipated by the security manager rather than the (possibly misinformed) adversary.

4.3 Cyber-physical attacker model

We model a cyber-physical attacker as an independent decision-making actor with the following properties.

Cyber and physical resources. We refer to an attacker's monetary budget, available human workforce and available ICT hardware as its physical resources. We refer to its ability to interact with the cyber sub-system of the power grid and affect its functionality as its cyber resources. Such cyber resources may include ICT software, knowledge of specific cyber vulnerabilities on the power grid cyber sub-system, rogue access points, *etc.*. In fine, the cyber and physical resources of an adversary jointly restrict the set of threat vectors it may launch against the electric power grid. We will represent such restriction by denoting the decision space of any cyber-attacker j as $\mathbf{a}^j \in \mathscr{A}^j \subseteq \mathscr{A}$.

Understanding of the cyber sub-system functionalities. In section 4.1 we introduced an abstract model of the power grid cyber sub-system through interface variable vectors $(\mathbf{y}, \mathbf{w})$ and equalities (4.1 – 4.2). Here, we wish to acknowledge that the distinct understanding of any cyber-attacker j on the organization and functionalities of the cyber sub-system leads to a distinct instance of this abstract model. For example, a cyber-attacker with insider priviledges should have better understanding of the cyber sub-system with respect to an outsider that has performed limited reconnaissance. This would ultimately translate in a more precise model of the cyber sub-system. We will add a tilde ($\tilde{}$) and superscript j to distinguish the cyber sub-system model belonging to a specific cyber-attacker

as,

$\forall \mathbf{a} \in \mathscr{A}^j \subseteq \mathscr{A}$:

$$\tilde{\mathbf{y}}^j(\mathbf{a}) = \tilde{\mathbf{y}}^j_{\emptyset} + \mathbf{d}\tilde{\mathbf{y}}^j(\tilde{\mathbf{m}}^j, \mathbf{a}), \tag{4.11}$$

$$\tilde{\mathbf{w}}^j(\mathbf{a}) = \tilde{\mathbf{w}}^j_{\emptyset} + \mathbf{d}\tilde{\mathbf{w}}^j(\tilde{\mathbf{m}}^j, \mathbf{a}). \tag{4.12}$$

Notice that in we intentionally use the tilde symbol and subscript both in functions $\mathbf{d}\tilde{\mathbf{y}}^j(\cdot,\cdot), \mathbf{d}\tilde{\mathbf{w}}^j(\cdot,\cdot)$ and their argument $\tilde{\mathbf{m}}^j$. Indeed, an adversary may have correct understanding of the cyber sub-system organization but incomplete/wrong understanding of the active cyber sub-system countermeasures (and vice versa).

Understanding of the electric power grid operation. We will model the power grid operational strategy in the style of (2.18), that is as the optimization problem of seeking to secure the system against a pre-defined set of contingencies, or at least minimizing the extent of insecurity. Further, we also acknowledge here that the adversary's information/understanding of the power grid operational strategy is not necessarily correct. For instance, an adversary may not know the precise objectives of real-time operation (*e.g.*, exact generation marginal cost values), the precise levels for the inequality constraints on the system state variables (*e.g.*, acceptable voltage limits, power flow values, *etc.*) and on decision variables (*e.g.*, generation ramping rates), and/or the precise physical model that the power grid operator would use to determine its optimal decisions. We once again employ the tilde and superscript j to distinguish the power grid operational strategy, as modeled by a specific cyber-attacker.

$$\begin{aligned} &\min_{\tilde{\mathbf{x}}^j, \tilde{\mathbf{u}}^j, \tilde{\mathbf{r}}^j} \tilde{f}^j_{\mathbf{ur}}(\tilde{\mathbf{u}}^j, \tilde{\mathbf{r}}^j, \tilde{\mathbf{y}}^j), \\ &\qquad \text{subject to:} \\ &\qquad \tilde{\mathbf{x}}^j \in \tilde{\mathscr{X}}^j_r(\tilde{\mathbf{u}}^j, \tilde{\mathbf{r}}^j, \tilde{\mathbf{y}}^j), \\ &\qquad \tilde{\mathbf{u}}^j \in \tilde{\mathscr{U}}^j(\tilde{\mathbf{p}}^j, \tilde{\mathbf{y}}^j), \\ &\qquad \tilde{\mathbf{r}}^j \in \tilde{\mathscr{R}}^j(\tilde{\mathbf{y}}^j). \end{aligned} \tag{4.13}$$

Motivation. As comprehensively reported in the previous deliverables of the CYPRESS project (see section 4.3.4 in [7] and 3.1.3 in [11]) adversaries may have a broad range of roles and motivations (from terrorist organizations to poorly trained employees). In the context of mathematical programming, we can describe the motivation of a cyber-attacker by means of (a) severity constraints on the state vector of the electric power grid and (b) an objective function to maximize through the choice of a threat vector.

Severity constraints on the power grid state vector can be used to describe the intended effect of a threat in technical terms. The amount of involuntary load shedding upon occurrence of an unwanted event is a very popular severity measure in the electric power systems literature. Adopting such measure, example severity constraints would express the minimum amount of involuntary load shedding that an attacker seeks to provoke. If such constraint cannot be satisfied by any attack vector, the attacker would consider the impact of its actions insignificant and would rather not launch an attack. An alternative example for severity constraints is provided in [21]. In this work, severity is measured in terms of the number of transmission branches that would sustain a significant (*i.e.*, above a certain threshold) overload in the aftermath of a cyber-physical attack. The respective severity constraint imposes a lower bound on the said number of branches, modeling an attacker that wishes to provoke a challenging instance of power grid insecurity.

Notice that to compute severity and ascertain the satisfaction of respective constraints, an adversary needs its own model of the grid physics. Its purpose is to simulate the power grid at a level of detail that matches its bespoke interests. Such level of detail may be more coarse-grained or more refined with respect to the physical model embedded in the operational strategies of the system operator. In our compact notation, it suffices to represent severity in function of the decisions taken

by the power grid operator $(\tilde{\mathbf{u}}^{j\star}, \tilde{\mathbf{r}}^{j\star})$ as well as the field-level interface variables $\tilde{\mathbf{w}}^j$. Indeed, given such vectors and a suitably chosen physical model, a cyber-attacker could compute the power grid state vector as well as evaluate its severity (*e.g.*, in case of a false data injection attack against the load demand measurements, a cyber-attacker could use the dispatching decisions of the power grid operator along with (her best guess) of the true demand values to solve the AC power flow equations and evaluate the extent of branch overloads). We will henceforth denote severity constraints as $\tilde{s}^j(\tilde{\mathbf{u}}^{j\star}, \tilde{\mathbf{r}}^{j\star}, \tilde{\mathbf{w}}^j(\mathbf{a}^j)) \geq \underline{\mathbf{s}}_{\mathbf{a}}^j$[1].

Objective function $f_{\mathbf{a}}^j(\tilde{\mathbf{u}}^{j\star}, \tilde{\mathbf{r}}^{j\star}, \tilde{\mathbf{w}}^j(\mathbf{a}^j), \mathbf{a}^j)$ accounts for the disruptive preferences of a malicious actor seeking to disrupt the functionality of the physical system (*e.g.*, seeking to maximize the extent of involuntary load shedding). We include the adversary's anticipation for the security management actions to be applied by the system operator $(\tilde{\mathbf{u}}^{j\star}, \tilde{\mathbf{r}}^{j\star})$ as a first and second argument in its objective function. It stands to account for the security management costs associated with responding to the threat instantiated by the adversary. Along with the 3rd argument, that is the field-level interface variables as perceived by the attacker, these two vectors can be combined to also monetize the severity of the impact of a cyber-physical threat on the electric power grid. The final argument is the threat vector, which in the general case may require the use of costly resources to be activated.

Complete cyber-physical attacker model. We put together a complete model for the decision making of a cyber-physical adversary by combining its cyber and physical resources, its understanding on the cyber sub-system, its knowledge of the electric power grid operation as well as its motivation in the following maximization problem.

$$\max_{\mathbf{a}^j} f_{\mathbf{a}}^j(\tilde{\mathbf{u}}^{j\star}, \tilde{\mathbf{r}}^{j\star}, \tilde{\mathbf{w}}^j(\mathbf{a}^j), \mathbf{a}^j), \tag{4.14}$$

subject to:

$$\mathbf{a}^j \in \mathscr{A}^j, \tag{4.15}$$

$$\tilde{s}^j(\tilde{\mathbf{u}}^{j\star}, \tilde{\mathbf{r}}^{j\star}, \tilde{\mathbf{w}}^j(\mathbf{a}^j)) \geq \underline{\mathbf{s}}_{\mathbf{a}}^j, \tag{4.16}$$

$$\tilde{\mathbf{y}}^j(\mathbf{a}^j) = \tilde{\mathbf{y}}_{\emptyset}^j + \mathbf{d}\tilde{\mathbf{y}}^j(\tilde{\mathbf{m}}^j, \mathbf{a}^j), \tag{4.17}$$

$$\tilde{\mathbf{w}}^j(\mathbf{a}^j) = \tilde{\mathbf{w}}_{\emptyset}^j + \mathbf{d}\tilde{\mathbf{w}}^j(\tilde{\mathbf{m}}^j, \mathbf{a}^j), \tag{4.18}$$

$$\tilde{\mathbf{u}}^{j\star}, \tilde{\mathbf{r}}^{j\star} \in \underset{\tilde{\mathbf{x}}^j, \tilde{\mathbf{u}}^j, \tilde{\mathbf{r}}^j}{\arg\min}\, \tilde{f}_{\mathbf{ur}}^j(\tilde{\mathbf{u}}^j, \tilde{\mathbf{r}}^j, \tilde{\mathbf{y}}^j(\mathbf{a}^j)), \tag{4.19}$$

subject to:

$$\tilde{\mathbf{x}}^j \in \tilde{\mathscr{X}}_r^j(\tilde{\mathbf{u}}^j, \tilde{\mathbf{r}}^j, \tilde{\mathbf{y}}^j(\mathbf{a}^j)), \tag{4.20}$$

$$\tilde{\mathbf{u}}^j \in \tilde{\mathscr{U}}^j(\tilde{\mathbf{p}}^j, \tilde{\mathbf{y}}^j(\mathbf{a}^j)), \tag{4.21}$$

$$\tilde{\mathbf{r}}^j \in \tilde{\mathscr{R}}^j(\tilde{\mathbf{y}}^j(\mathbf{a}^j)). \tag{4.22}$$

Problem (4.14 – 4.22) is a quite generic model for the interaction between an adversary and the operator of the electric power grid. It can be declined to instantiate cases corresponding to the interaction between a specific adversary and the power grid operator, by specifying the precise mathematical formulas for all functions as well as some constants symbolically represented here. As a proof-of-concept, appendix A presents the declination of this formulation in the case of load redistribution cyber-physical attacks. Section A.3 in particular provides an explicit example of the precise formulation of all functions appearing in (4.14 – 4.22) as well as all relevant constants. In the model of section A.3:

- objective function (4.14) is stated as the summation of the magnitude of the branch overloads induced by a cyber-physical attack.

[1] We use the tilde ˜ to denote that severity is evaluated with the (possibly inaccurate) models and data available to an exogenous adversary rather than the (more accurate) models and data available to the power system operator/planner, and superscript j to specify this adversary.

- Resource constraints (4.15) are stated as a limit on the number of load meters that can be simultaneously manipulated by a cyber-physical attacker. Along with undetectability constraints, based on the attacker's understanding of the cyber sub-system, these in fine restrict the cyber sub-system state modifications that the cyber-attacker may induce.

- Severity constraints (4.16) are composed of the DC approximation of the power flow equations as a physical model, as well as a group of inequality constraints to only account for the branches that sustain "significant" overloads above a certain threshold.

- The attacker's anticipated effect on the operation-level field variable vector (4.17) is modeled by an additional demand term in the nodal power balance constraint of the optimization problem describing the behavior of the real-time operator.

- The attacker's anticipated effect on the field-level variable vector 4.18 is implicitly represented, by maintaining the original demand values in the power flow model used to evaluate the severity of the attack.

- The model of the real-time operator (4.19 – 4.22) is stated as a DC Optimal Power Flow problem seeking to minimize the system generation cost.

- Both the power flow equations used to evaluate the severity constraints (4.16) and the model of the real-time operator include imprecise reactance values.

4.4 Cyber-physical operation planner (a.k.a. security manager) decision-making problem

We hereafter focus on the decision-making problem of the so-called integrated cyber-physical operation planner seeking to secure the operation of the electric power system while (a) acting both on the cyber sub-system and on the physical power system and (b) anticipating the adversarial decisions of cyber-attackers as well as the resulting reactions of the power system real-time operator.

4.4.1 Deterministic setting: facing a single attacker profile

For ease of exposition, we begin by stating the decision-making problem of a planner facing a single cyber-attacker j whose properties are known with certainty as shown in (4.23 – 4.43). The coding of all equations corresponds to the different colors used to illustrate the different decision makers in Figures 4.1 and 4.2. We start by a model of an attacker that is reasonably well informed and then present two variants where the attacker is respectively less well or much better informed.

Facing a reasonably well informed attacker

We first consider the case where the cyber-physical model of the attacker may differ from the one used by the security planner, while on the other hand we assumer that the attacker is able to observe the decisions take by the security planner.

The top level (4.23 – 4.30) corresponds to the perspective of the planner on the cyber-physical electric power grid. As per objective function (4.23) we consider a planner who seeks to jointly minimize the cost of physical planning actions (*e.g.*, starting/stopping generating units, accepting/cancelling maintenance requests *etc.*), cyber sub-system countermeasures (*e.g.*, firewalls, encrypted communications,*etc.*) along with the cost of operating the system in the occassion that no cyber-attack materializes. Constraint (4.24) denotes the decision space for the planning decisions acting on the physical power grid while, as introduced in chapter 2 constraints (4.25 – 4.26) express that the system should be operable with acceptable security in case no cyber-physical attack happens. Notice

symbol $\hat{\mathbf{y}}$ appearing in both these constraint expressions to represent the expected value of the so-called operational-level interface variable. Provided that the cyber sub-system is well functioning, this parameter is expected to accurately reflect the power grid operational condition.

Constraint (4.27) denotes the decision space for the countermeasures acting on the cyber sub-system of the power grid. Equalities (4.28 – 4.29) express the planner's anticipation for the effect of the cyber-attacker's decisions on the performance of the cyber sub-system of the electric power grid. These equations symbolically express that the operational-level and field-level interface variables may deviate from the respective expected values according to the cyber-attacker's optimal decision strategy (shown as $\mathbf{a}^{j\star}$). Finally, constraint (4.30) defines the sought level of cyber-physical security with respect to threat posed by the cyber-physical adversary j. We symbolically express this here as a restriction on the potential severity of the resulting power grid state and while also taking into account the decisions that the real-time operator would take following the realization of the cyber-attack ($\mathbf{u}^{j\star}, \mathbf{r}^{j\star}$). For example, this restriction can be an upper bound on the branch overloads, over/-under- voltages, emergency load shedding extent *etc.*. Notice that we include here the field-level interface variable as an argument of the severity expression.

The middle level (4.31 – 4.39) expresses the planner's perspective on the anticipated behavior of the cyber-physical adversary. The detailed statement of this problem is ellaborated in the precedent Section 4.3. The lower level (4.40 – 4.43) is the anticipated reaction of the real-time operator both to the planning decisions of the integrated planner as well as the threat vector of the cyber-physical attacker.

$$\min_{\mathbf{p},\mathbf{m},\hat{\mathbf{x}},\hat{\mathbf{u}}} \left\{ f_p(\mathbf{p}) + f_m(\mathbf{m}) + f_{\mathbf{u}}(\hat{\mathbf{u}},\hat{\mathbf{y}}) \right\}, \tag{4.23}$$

subject to:

$$\mathbf{p} \in \mathscr{P}, \tag{4.24}$$

$$\hat{\mathbf{x}} \in \mathscr{X}(\hat{\mathbf{u}},\hat{\mathbf{y}}), \tag{4.25}$$

$$\hat{\mathbf{u}} \in \mathscr{U}(\mathbf{p},\hat{\mathbf{y}}), \tag{4.26}$$

$$\mathbf{m} \in \mathscr{M}, \tag{4.27}$$

$$\mathbf{y}(\mathbf{a}^{j\star}) = \hat{\mathbf{y}} + \mathbf{dy}(\mathbf{m},\mathbf{a}^{j\star}), \tag{4.28}$$

$$\mathbf{w}(\mathbf{a}^{j\star}) = \hat{\mathbf{w}} + \mathbf{dw}(\mathbf{m},\mathbf{a}^{j\star}), \tag{4.29}$$

$$s(\mathbf{u}^{j\star},\mathbf{r}^{j\star},\mathbf{w}^{j}(\mathbf{a}^{j\star})) \leq \bar{\mathbf{s}}_{\mathbf{p}}, \tag{4.30}$$

$$\mathbf{a}^{j\star} \in \underset{\mathbf{a}^j}{\arg\max}\, f_{\mathbf{a}}^{j}(\tilde{\mathbf{u}}^{j\star},\tilde{\mathbf{r}}^{j\star},\tilde{\mathbf{w}}^{j}(\mathbf{a}^{j}),\mathbf{a}^{j}), \tag{4.31}$$

subject to:

$$\mathbf{a}^{j} \in \mathscr{A}^{j}, \tag{4.32}$$

$$\tilde{s}^{j}(\tilde{\mathbf{u}}^{j\star},\tilde{\mathbf{r}}^{j\star},\tilde{\mathbf{w}}^{j}(\mathbf{a}^{j})) \geq \underline{\mathbf{s}}_{\mathbf{a}}^{j}, \tag{4.33}$$

$$\tilde{\mathbf{y}}^{j}(\mathbf{a}^{j}) = \tilde{\mathbf{y}}_{\emptyset}^{j} + \mathbf{d}\tilde{\mathbf{y}}^{j}(\mathbf{m},\mathbf{a}^{j}), \tag{4.34}$$

$$\tilde{\mathbf{w}}^{j}(\mathbf{a}^{j}) = \tilde{\mathbf{w}}_{\emptyset}^{j} + \mathbf{d}\tilde{\mathbf{w}}^{j}(\mathbf{m},\mathbf{a}^{j}), \tag{4.35}$$

$$\tilde{\mathbf{u}}^{j\star},\tilde{\mathbf{r}}^{j\star} \in \underset{\tilde{\mathbf{x}}^j,\tilde{\mathbf{u}}^j,\tilde{\mathbf{r}}^j}{\arg\min}\, \tilde{f}_{\mathbf{ur}}^{j}(\tilde{\mathbf{u}}^{j},\tilde{\mathbf{r}}^{j},\tilde{\mathbf{y}}^{j}(\mathbf{a}^{j})), \tag{4.36}$$

subject to:

$$\tilde{\mathbf{x}}^{j} \in \tilde{\mathscr{X}}_{r}^{j}(\tilde{\mathbf{u}}^{j},\tilde{\mathbf{r}}^{j},\tilde{\mathbf{y}}^{j}(\mathbf{a}^{j})), \tag{4.37}$$

$$\tilde{\mathbf{u}}^{j} \in \tilde{\mathscr{U}}^{j}(\mathbf{p},\tilde{\mathbf{y}}^{j}(\mathbf{a}^{j})), \tag{4.38}$$

$$\tilde{\mathbf{r}}^{j} \in \tilde{\mathscr{R}}^{j}(\tilde{\mathbf{y}}^{j}(\mathbf{a}^{j})), \tag{4.39}$$

$$\mathbf{u}^{j\star},\mathbf{r}^{j\star} \in \underset{\mathbf{x}^j,\mathbf{u}^j,\mathbf{r}^j}{\arg\min}\, f_{\mathbf{ur}}^{j}(\mathbf{u}^{j},\mathbf{r}^{j},\tilde{\mathbf{y}}^{j}(\mathbf{a}^{j\star})), \tag{4.40}$$

subject to:

$$\mathbf{x}^{j} \in \mathscr{X}^{j}(\mathbf{u}^{j},\mathbf{r}^{j},\mathbf{y}^{j}(\mathbf{a}^{j\star})), \tag{4.41}$$

$$\mathbf{u}^{j} \in \mathscr{U}(\mathbf{p},\mathbf{y}^{j}(\mathbf{a}^{j\star})), \tag{4.42}$$

$$\mathbf{r}^{j} \in \mathscr{R}(\mathbf{y}^{j}(\mathbf{a}^{j\star})). \tag{4.43}$$

Facing a less well informed attacker

Considering the case wherein attacker j has no ability to observe the planner's decisions but rather assumes the values for $(\tilde{\mathbf{p}}^j, \tilde{\mathbf{m}}^j)$ independently, the threat vector $\mathbf{a}^{j\star}$ can be pre-defined from the solution of (4.14 – 4.22). In such case, it becomes a parameter for the decision making problem of the planner, which in turn reduces to the bi-level problem (4.44 – 4.55). Notice that the decision-making actors explicitly modeled in (4.44 – 4.55) are the planner and real-time operator of the electric power grid, that are by default cooperative. Under carefully-chosen assumptions on the types of actions available to these actors as well as the precise models representing the physical power grid and its cyber sub-system, instances of (4.44 – 4.55) are solvable by robust optimization techniques already explored in the power systems security management literature [38].

$$\min_{\mathbf{p},\mathbf{m},\hat{\mathbf{x}},\hat{\mathbf{u}}} \Big\{ f_p(\mathbf{p}) + f_m(\mathbf{m}) + f_{\mathbf{u}}(\hat{\mathbf{u}},\hat{\mathbf{y}}) \Big\}, \tag{4.44}$$

subject to:

$$\mathbf{p} \in \mathscr{P}, \tag{4.45}$$

$$\hat{\mathbf{x}} \in \mathscr{X}(\hat{\mathbf{u}},\hat{\mathbf{y}}), \tag{4.46}$$

$$\hat{\mathbf{u}} \in \mathscr{U}(\mathbf{p},\hat{\mathbf{y}}), \tag{4.47}$$

$$\mathbf{m} \in \mathscr{M}, \tag{4.48}$$

$$\mathbf{y}(\mathbf{a}^{j\star}) = \hat{\mathbf{y}} + \mathbf{dy}(\mathbf{m},\mathbf{a}^{j\star}), \tag{4.49}$$

$$\mathbf{w}(\mathbf{a}^{j\star}) = \hat{\mathbf{w}} + \mathbf{dw}(\mathbf{m},\mathbf{a}^{j\star}), \tag{4.50}$$

$$s(\mathbf{u}^{j\star},\mathbf{r}^{j\star},\mathbf{w}^j(\mathbf{a}^{j\star})) \leq \bar{\mathbf{s}}_{\mathbf{p}}, \tag{4.51}$$

$$\mathbf{u}^{j\star},\mathbf{r}^{j\star} \in \underset{\mathbf{x}^j,\mathbf{u}^j,\mathbf{r}^j}{\arg\min}\, f^j_{\mathbf{ur}}(\mathbf{u}^j,\mathbf{r}^j,\tilde{\mathbf{y}}^j(\mathbf{a}^{j\star})), \tag{4.52}$$

subject to:

$$\mathbf{x}^j \in \mathscr{X}^j_r(\mathbf{u}^j,\mathbf{r}^j,\mathbf{y}^j(\mathbf{a}^{j\star})), \tag{4.53}$$

$$\mathbf{u}^j \in \mathscr{U}^j(\mathbf{p}^j,\mathbf{y}^j(\mathbf{a}^{j\star})), \tag{4.54}$$

$$\mathbf{r}^j \in \mathscr{R}^j(\mathbf{y}^j(\mathbf{a}^{j\star})). \tag{4.55}$$

Facing a perfectly well informed attacker

Alternatively, assuming that the adversary has perfect information on the cyber-physical system, as well as the ability to correctly observe the planners decisions there is no need to distinguish between the cyber-attacker's model of the system and the models that are available to the planner and operator. In such case, the original problem (4.23 – 4.43) reduces to the tri-level optimization problem (4.56 – 4.70). Again, depending on the specific choices for the types of actions available to these actors as well as the precise models representing the physical power grid and its cyber sub-system, instances of (4.56 – 4.43) are solvable by trilevel optimization techniques already explored in the power systems security management literature [39].

$$\min_{\mathbf{p},\mathbf{m},\hat{\mathbf{x}},\hat{\mathbf{u}}} \Big\{ f_p(\mathbf{p}) + f_m(\mathbf{m}) + f_{\mathbf{u}}(\hat{\mathbf{u}},\hat{\mathbf{y}}) \Big\}, \tag{4.56}$$

subject to:

$$\mathbf{p} \in \mathscr{P}, \tag{4.57}$$

$$\hat{\mathbf{x}} \in \mathscr{X}(\hat{\mathbf{u}},\hat{\mathbf{y}}), \tag{4.58}$$

$$\hat{\mathbf{u}} \in \mathscr{U}(\mathbf{p},\hat{\mathbf{y}}), \tag{4.59}$$

$$\mathbf{m} \in \mathscr{M}, \tag{4.60}$$

$$s(\mathbf{u}^{j\star},\mathbf{r}^{j\star},\mathbf{w}^j(\mathbf{a}^{j\star})) \leq \bar{\mathbf{s}}_{\mathbf{p}}, \tag{4.61}$$

$$\mathbf{a}^{j\star} \in \arg\max_{\mathbf{a}^j} f^j_{\mathbf{a}}(\mathbf{u}^{j\star},\mathbf{r}^{j\star},\mathbf{w}(\mathbf{a}^j),\mathbf{a}^j), \tag{4.62}$$

subject to:

$$\mathbf{a}^j \in \mathscr{A}^j, \tag{4.63}$$

$$s^j(\mathbf{u}^{j\star},\mathbf{r}^{j\star},\mathbf{w}(\mathbf{a}^j)) \geq \underline{\mathbf{s}}^j_{\mathbf{a}}, \tag{4.64}$$

$$\mathbf{y}(\mathbf{a}^j) = \mathbf{y}_\emptyset + \mathbf{dy}(\mathbf{m},\mathbf{a}^j), \tag{4.65}$$

$$\mathbf{w}(\mathbf{a}^j) = \mathbf{w}_\emptyset + \mathbf{dw}(\mathbf{m},\mathbf{a}^j), \tag{4.66}$$

$$\mathbf{u}^{j\star},\mathbf{r}^{j\star} \in \arg\min_{\mathbf{x}^j,\mathbf{u}^j,\mathbf{r}^j} f_{\mathbf{ur}}(\mathbf{u}^j,\mathbf{r}^j,\mathbf{y}^j(\mathbf{a}^j)), \tag{4.67}$$

subject to:

$$\mathbf{x}^j \in \mathscr{X}_r(\mathbf{u}^j,\mathbf{r}^j,\mathbf{y}(\mathbf{a}^j)), \tag{4.68}$$

$$\mathbf{u}^j \in \mathscr{U}^j(\mathbf{p},\mathbf{y}(\mathbf{a}^j)), \tag{4.69}$$

$$\mathbf{r}^j \in \mathscr{R}^j(\mathbf{y}(\mathbf{a}^j)). \tag{4.70}$$

4.4.2 Stochastic variants

As with any cyber-physical security management application, identifying the precise properties of a cyber-attacker that may launch a threat vector against the power grid is obviously a non-trivial task. In real-life applications, intelligence gathering as well as vulnerability analysis are the main tools used to define a credible set of potential cyber-attackers, differentiated in terms of their motivations, strategies, resources and capabilities. Here, we take the approach of portraying the inevitable uncertainty of a cyber-physical planner on the precise properties of a cyber-attacker that may wish to disrupt the power grid functionality by means of a discrete set $j \in \mathscr{J}$.

Alternative credible cyber-attackers $j \in \mathscr{J}$ are differentiated in terms of the threat vectors they are able to launch (a.ka. their cyber and physical resources in the parlance of section 4.3), their motivation (as expressed by an optimization objective and possibly a severity constraint) as well as their understanding of the cyber-physical system they seek to disrupt (mathematically expressed by a group of equality/inequality constraints). Reference [12] provides an example of how one may populate such a set on the basis of cyber sub-system and power grid models and while using the concept of *bounded rationality*. This work exploits the so-called cognitive hierarchy framework [40] to also assign each member of set $\mathscr{J}$ with a probability value, corresponding to the degree of belief that such an agent may indeed exist and consider attacking the cyber-physical electric power grid. We therefore assume that the models for a set of credible cyber-attackers $\mathscr{J}$ along with their respective occurrence probabilities $\pi_j : \sum_{j\in\mathscr{J}} = 1 - \pi_\emptyset$ is available as input to the cyber-physical security management decision-making problem.

Minimizing monetized severity in expectation

Given such input, a first approach to stochastic cyber-physical security management could be to trade-off the costs of the enhanced cyber and physical measures against the expected consequences of cyber-physical insecurity by monetizing the severity metric. Mathematically such an approach can be stated as in (4.71 – 4.76). The last term in objective function (4.71) is the expected value of the monetized severity metric, over the alternative credible cyber-attackers. We argue that this approach is not well-suited for the problem in question. Solving (4.71 – 4.76) to optimality is only realistically plausible while replacing the symbolic objective/constraint expressions of the different actors with simple (mostly linear, convex and static) mathematical models. The most challenging properties of the power grid cyber-physical security management problem, such as fast temporal dynamics, non-convexities and non-linearities, have therefore to be ommited by default. Further, even though a solution may be optimal in tems of the cost *vs* expected benefit trade-off it still leaves the system (and its end-users) exposed to the risk of low-likelihood high impact catastrophic events. Notice that in the spirit of economic reasoning, we did not include the so-called severity constraint (4.30) limiting the technical extent of the potential threat impact on the power-grid functionality. In practice, such a constraint is a necessary proxy for avoiding unwanted phenomena, as well as for ensuring that the system is operated within the validity domain of the models used to identify optimal decisions.

$$\min_{\mathbf{p},\mathbf{m},\hat{\mathbf{x}},\hat{\mathbf{u}},\mathbf{s}} \left\{ f_p(\mathbf{p}) + f_m(\mathbf{m}) + f_{\mathbf{u}}(\hat{\mathbf{u}},\hat{\mathbf{y}}) + \mathbb{E}_{j\in\mathscr{J}} \left\{ f_{\mathbf{s}}\left(s(\mathbf{u}^{j\star},\mathbf{r}^{j\star},\mathbf{w}^{j}(\mathbf{a}^{j\star})) \right) \right\} \right\}, \tag{4.71}$$

subject to:

$$\mathbf{p} \in \mathscr{P}, \tag{4.72}$$

$$\hat{\mathbf{x}} \in \mathscr{X}(\hat{\mathbf{u}},\hat{\mathbf{y}}), \tag{4.73}$$

$$\hat{\mathbf{u}} \in \mathscr{U}(\mathbf{p},\hat{\mathbf{y}}), \tag{4.74}$$

$$\mathbf{m} \in \mathscr{M}, \tag{4.75}$$

subject to (4.28,4.29, 4.31 – 4.43)$\forall\, j \in \mathscr{J}$. (4.76)

Upper bounding severity in expectation

An alternative approach to stochastic cyber-physical power grid security management could be formulated by seeking to ensure that the expected severity of cyber-physical attacks may not exceed a certain level. Mathematically such an approach can be stated as in (4.77 – 4.83). Algorithmically such an approach is more advantageous with respect to (4.71 – 4.76) since ommiting the expectation term which involves middle and lower level decisions from objective function (4.77). Several column and constraint generation schemes can be leveraged towards the iterative satisfaction of constraint (4.77). Further, it is in principle possible to rely both on detailed analytical models as well as machine-learnt approximations to progressively evaluate the left-hand-side of (4.82) within such solution schemes. In other words, we believe that it is more viable to attempt to represent the (extreme) technical compexity of the issue at hand while solving a problem of the format (4.77 – 4.83) with respect to (4.71 – 4.76). Besides algorithmic solvability, it is also worth commenting on constraint (4.82) from a security management perspective. While it ensures that in expectation the sought severity level will not be exceeded (*e.g.,* in expectation the load that may have to be shed after any cyber-physical attack should be limited), it may turn out that cyber-physical attack event(s) exceeding such level are relatively probable.

$$\min_{\mathbf{p},\mathbf{m},\hat{\mathbf{x}},\hat{\mathbf{u}},\mathbf{s}} \left\{ f_p(\mathbf{p}) + f_m(\mathbf{m}) + f_{\mathbf{u}}(\hat{\mathbf{u}},\hat{\mathbf{y}}) \right\}, \tag{4.77}$$

subject to:

$$\mathbf{p} \in \mathscr{P}, \tag{4.78}$$

$$\hat{\mathbf{x}} \in \mathscr{X}(\hat{\mathbf{u}},\hat{\mathbf{y}}), \tag{4.79}$$

$$\hat{\mathbf{u}} \in \mathscr{U}(\mathbf{p},\hat{\mathbf{y}}), \tag{4.80}$$

$$\mathbf{m} \in \mathscr{M}, \tag{4.81}$$

$$\mathbb{E}_{j\in\mathscr{J}}\{s(\mathbf{u}^{j\star},\mathbf{r}^{j\star},\mathbf{w}^{j}(\mathbf{a}^{j\star}))\} \leq \bar{\mathbf{s}}_p, \tag{4.82}$$

subject to (4.28,4.29, 4.31 – 4.43)$\forall\, j \in \mathscr{J}$. (4.83)

Ensuring a small enough probability of too high cyber-attack severity

The alternative approach formulated in (4.84 – 4.90) rather seeks to ensure that the probability of cyber-physical attack event(s) exceeding the sought severity level is limited. In other words, it seeks to ensure that the power grid functionality will remain unaffacted by cyber-physical threats with high enough probability. From an algorithmic solvability perspective, we argue that it has similiar properties as (4.77 – 4.83) concerning the potential to rely on detailed analytical models and machine-learnt approximations within iterative constraint satisfaction schemes. We further believe that it is more comprehensive from a security management perspective, since providing an explicit guarantee that the chance of suffering unwanted consequences is as low as desired.

$$\min_{\mathbf{p},\mathbf{m},\hat{\mathbf{x}},\hat{\mathbf{u}},\mathbf{s}} \left\{ f_p(\mathbf{p}) + f_m(\mathbf{m}) + f_{\mathbf{u}}(\hat{\mathbf{u}},\hat{\mathbf{y}}) \right\}, \tag{4.84}$$

subject to:

$$\mathbf{p} \in \mathscr{P}, \tag{4.85}$$

$$\hat{\mathbf{x}} \in \mathscr{X}(\hat{\mathbf{u}},\hat{\mathbf{y}}), \tag{4.86}$$

$$\hat{\mathbf{u}} \in \mathscr{U}(\mathbf{p},\hat{\mathbf{y}}), \tag{4.87}$$

$$\mathbf{m} \in \mathscr{M}, \tag{4.88}$$

$$\mathbb{P}_{j\in\mathscr{J}}\left\{ s(\mathbf{u}^{j\star},\mathbf{r}^{j\star},\mathbf{w}^{j}(\mathbf{a}^{j\star})) \leq \bar{\mathbf{s}}_p \right\} \geq 1-\epsilon, \tag{4.89}$$

subject to (4.28,4.29, 4.31 – 4.43)$\forall\, j \in \mathscr{J}$. (4.90)

5

Conclusions

This document investigates the stakes of extending the multi-stage stochastic programming approach from the (physical) electric power grid security management context to the cyber-physical electric power grid security management context.

In the context of (physical) electric power grid security, preventive security management means taking decisions before any credible contingency event happens so as to ensure that the system can withstand this event without any further intervention from its operator. In other words, preventive measures (*e.g.*, choosing the transmission grid topology, the dispatching of power generating units, the settings of automated protection devices) are put in place so as to establish a certain confidence that the power grid can "safely absorb" any credible contingency event. Extending this approach from the domain of physical security to the domain of cyber-physical security entails:

1. ensuring the performance of the cyber sub-system, through protection and detection & mitigation countermeasures, so that it may not induce any event that cannot be "safely absorbed" by the power grid without any further intervention from its operator, and/or,

2. adapting the physical controls of the electric power grid so that it can "safely absorb" a broader set of contingencies, including events induced by malfunctions of the cyber sub-system.

Further, in the context of (physical) electric power grid security, preventive/corrective security management means taking decisions before and/or after any credible contingency event happens. Actions before a contingency event aim to ensure that the system will not degrade severely before its operator has an opportunity to respond. Corrective actions are only to be applied after the event and restore the full functionality of the electric power grid.

Extending this approach from the domain of physical security to the domain of cyber-physical security entails:

1. ensuring the performance of the cyber sub-system so that it may not induce any contingency event that cannot be promptly addressed by the operator of the electric power grid, and/or,

2. adapting the physical controls of the electric power grid to be able to withstand a broader set of credible events, including contingencies induced by malfunctions of the cyber sub-system.

While the verbal redefinition of the scope for power system preventive and corrective controls is easy to establish, the challenges of extending the current security management approach in order to be able to cope with cyber-physical threats should not be underestimated. We wish to emphasize on (i) the need to also model the cyber sub-system of the power grid at a suitable level of detail that allows to model both its vulnerabilities and the countermeasures that may be put in place against these, (ii) the fact that most relevant cyber-physical threats are not exogenous *fatal* eventualities by rather intentional malicious attacks designed by intelligent agents and (iii) the complexity of the resulting multi-actor decision making problem concerning the integrated cyber-physical power grid. The contents of this report portray these fundamental challenges.

First, chapter 3 reviews and classifies the threats and countermeasures that may act on the cyber sub-system of the electric power grid. The classification of cyber-physical threats into three main categories is relevant for considering the so-called *bigger-picture*, however one cannot fail to notice that each distinct threat may exploit a different functionality of the cyber sub-system and/or target a different functionality of the physical power grid. This could in turn translate into completely different modeling requirements of the cyber sub-system for different events simultaneously threatening the power system operation. Likewise, while it is possible to group together countermeasures in terms of their cyber-security purpose into protection, detection and mitigation schemes, there is quite some diversity of actual countermeasures acting on specific functionalities of the power grid cyber sub-system. This diversity in threats and countermeasures typically results in the development of bespoke solutions for distinct threats rather than a holistic approach to managing the risk of the system. To circumvent this challenge, we relied here on the cyber sub-system modeling abstraction introduced in section 4.1. It focuses on the effect of cyber threats on the so-called interface variables between the physical components, the cyber sub-system as well as the human operators (and control automata) of the power grid. While it allows us to discuss on the principles of a holistic security management approach, irrespectively of specific threat and countermeasure properties, the complexity of modeling the actual cyber sub-system seems inevitable for practical cyber-physical security management.

The fact that cyber-physical attacks are not random events but well designed actions of malicious adversaries is a complication that comes on top of the precedent challenge. Here, it is important to notice that such actors are in general bounded-rational thinkers, with self-determined motivations as well as imperfect information. It seems inconceivable that one may know in advance the precise properties of the malicious adversary that will eventually target to disrupt a cyber-physical system, so as to put in place the preventive measures that would protect the system from the actions of such agent. Taking a robust perspective, predefining a worst-case adversary against whom the system preventive measures are designed is a possible way out. However, this typically comes at increased cost, which is known as *the price of conservativeness*. In section (4.3) we built on top the idea presented in [12] to portray the uncertainty on the adversary characteristics by defining a set of alternative cyber-attackers, differentiated by their motivations, resources, understanding and modeling skills on the cyber-physical electric power system. As an example, in appendix A we differentiated alternative cyber-attackers in terms of the accuracy of the technical data they possess on the electric power grid. Assimilating the different cyber-attacker models already available in the literature could be a first approach for building such a set of alternative cyber-attackers to serve as input for cyber-physical security management.

The aforementioned challenges combine to make the mathematical statement and algorithmic resolution of cyber-physical power system security management stochastic decision-making problems a cumbersome excercise. In section 4.4 of this report we adopted the perspective of an integrated cyber-physical power grid planner, simultaneously taking preventive decisions both on the

cyber sub-system and physical power grid and while relying on the power grid operator for *physical-only* corrective controls. We considered alternative structures for the constraints and objective of such problem and briefly discussed the merits of each alternative both from a risk management perspective and from an algorithmic solvability perspective. The continuation of the CYPRESS WP3 research effort concerns both the precise mathematical models that should be used to formulate relevant instances of these stochastic problems as well as the development of proof-of-concept solution approaches.

Bibliography

[1] E. Karangelos and L. Wehenkel, "Post-contingency corrective control failure: A risk to neglect or a risk to control?" In *2018 IEEE International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, 2018, pp. 1–6. DOI: 10.1109/PMAPS.2018.8440348.

[2] H. He and J. Yan, "Cyber-physical attacks and defences in the smart grid: A survey," *IET Cyber-Physical Systems: Theory & Applications*, vol. 1, no. 1, pp. 13–27, 2016. DOI: https://doi.org/10.1049/iet-cps.2016.0019. [Online]. Available: https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/iet-cps.2016.0019.

[3] Z. E. Mrabet, N. Kaabouch, H. E. Ghazi, and H. E. Ghazi, "Cyber-security in smart grid: Survey and challenges," *Computers & Electrical Engineering*, vol. 67, pp. 469–482, 2018, ISSN: 0045-7906. DOI: https://doi.org/10.1016/j.compeleceng.2018.01.015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0045790617313423.

[4] A. Gusrialdi and Z. Qu, "Smart grid security: Attacks and defenses," in *Smart Grid Control: Overview and Research Opportunities*, J. Stoustrup, A. Annaswamy, A. Chakrabortty, and Z. Qu, Eds. Cham: Springer International Publishing, 2019, pp. 199–223, ISBN: 978-3-319-98310-3.

[5] H. Zhang, B. Liu, and H. Wu, "Smart grid cyber-physical attack and defense: A review," *IEEE Access*, vol. 9, pp. 29 641–29 659, 2021. DOI: 10.1109/ACCESS.2021.3058628.

[6] I. Zografopoulos, J. Ospina, X. Liu, and C. Konstantinou, "Cyber-physical energy systems security: Threat modeling, risk assessment, resources, metrics, and case studies," *IEEE Access*, vol. 9, pp. 29 775–29 818, 2021. DOI: 10.1109/ACCESS.2021.3058403.

[7] E. Karangelos, K. Thoelen, F. Faghihi, *et al.*, "Report D1.1: Describing the selected performance metrics," The CYPRESS project, Tech. Rep., 2021.

[8] F. Capitanescu, J. Martinez Ramos, P. Panciatici, *et al.*, "State-of-the-art, challenges, and future trends in security constrained optimal power flow," *Electric Power Systems Research*, vol. 81, no. 8, pp. 1731–1741, 2011, ISSN: 0378-7796. DOI: https://doi.org/10.1016/j.epsr.2011.04.003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378779611000885.

[9] F. Capitanescu, "Critical review of recent advances and further developments needed in AC optimal power flow," *Electric Power Systems Research*, vol. 136, pp. 57–68, 2016, ISSN: 0378-7796. DOI: https://doi.org/10.1016/j.epsr.2016.02.008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378779616300141.

[10] L. Roald, D. Pozo, A. Papavasiliou, D. K. Molzahn, J. Kazempour, and A. J. Conejo, "Power systems optimization under uncertainty: A review of methods and applications," *Electric Power Systems Research*, 2022.

[11] S. Ben Mariem, V. Rosseto, V. Guler, A. Godfraind, L. Mathy, and H. Ergun, "Report D1.2: Describing the relevant cyber-components, threats and barriers," The CYPRESS project, Tech. Rep., 2022.

[12] A. Sanjab and W. Saad, "On bounded rationality in cyber-physical systems security: Game-theoretic analysis with application to smart grid protection," in *2016 Joint Workshop on Cyber- Physical Security and Resilience in Smart Grids (CPSR-SG)*, 2016, pp. 1–6. DOI: 10.1109/CPSRSG.2016.7684101.

[13] M. A. Rahman and H. Mohsenian-Rad, "False data injection attacks with incomplete information against smart power grids," in *2012 IEEE Global Communications Conference (GLOBECOM)*, 2012, pp. 3153–3158. DOI: 10.1109/GLOCOM.2012.6503599.

[14] J. Zhang, Z. Chu, L. Sankar, and O. Kosut, "Can attackers with limited information exploit historical data to mount successful false data injection attacks on power systems?" *IEEE Transactions on Power Systems*, vol. 33, no. 5, pp. 4775–4786, 2018. DOI: 10.1109/TPWRS.2018.2818746.

[15] M. Potvin, "The AURORA vulnerability: The sword of Damocles over the heads of rotating machines," in *CIGRE Canada Conference*, 2019.

[16] A. Ashok, P. Wang, M. Brown, and M. Govindarasu, "Experimental evaluation of cyber attacks on automatic generation control using a CPS security testbed," in *2015 IEEE Power & Energy Society General Meeting*, 2015, pp. 1–5. DOI: 10.1109/PESGM.2015.7286615.

[17] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," vol. 14, no. 1, 2011-06, ISSN: 1094-9224. DOI: 10.1145/1952982.1952995. [Online]. Available: https://doi.org/10.1145/1952982.1952995.

[18] Y. Yuan, Z. Li, and K. Ren, "Modeling load redistribution attacks in power systems," *IEEE Transactions on Smart Grid*, vol. 2, no. 2, pp. 382–390, 2011. DOI: 10.1109/TSG.2011.2123925.

[19] J. Liang, L. Sankar, and O. Kosut, "Vulnerability analysis and consequences of false data injection attack on power system state estimation," *IEEE Transactions on Power Systems*, vol. 31, no. 5, pp. 3864–3872, 2015.

[20] J. Zhang and L. Sankar, "Physical system consequences of unobservable state-and-topology cyber-physical attacks," *IEEE Transactions on Smart Grid*, vol. 7, no. 4, 2016.

[21] E. Karangelos and L. Wehenkel, "Cyber–physical risk modeling with imperfect cyber-attackers," *Electric Power Systems Research*, vol. 211, p. 108 437, 2022, ISSN: 0378-7796. DOI: https://doi.org/10.1016/j.epsr.2022.108437.

[22] "Lessons learned – risks posed by Firewall Firmware Utilities," North American Reliability Council (NERC), Tech. Rep., 2019.

[23] Z. Li, M. Shahidehpour, A. Alabdulwahab, and A. Abusorrah, "Bilevel model for analyzing coordinated cyber-physical attacks on power systems," *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2260–2272, 2016. DOI: 10.1109/TSG.2015.2456107.

[24] Y. Xiang, L. Wang, and N. Liu, "Coordinated attacks on electric power systems in a cyber-physical environment," *Electric Power Systems Research*, vol. 149, pp. 156–168, 2017, ISSN: 0378-7796.

[25] H.-M. Chung, W.-T. Li, C. Yuen, W.-H. Chung, Y. Zhang, and C.-K. Wen, "Local cyber-physical attack for masking line outage and topology attack in smart grid," *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 4577–4588, 2019. DOI: 10.1109/TSG.2018.2865316.

[26] J. E. Sullivan and D. Kamensky, "How cyber-attacks in Ukraine show the vulnerability of the U.S. power grid," *The Electricity Journal*, vol. 30, no. 3, pp. 30–35, 2017, ISSN: 1040-6190.

[27] J. Slowik, "Crashoverride: Reassessing the 2016 Ukraine electric power event as a protection-focused attack," Dragos, Inc, Tech. Rep., 2019.

[28] A. Qiangsheng and L. Shi, "A game-theoretic analysis of cyber attack-mitigation in centralized feeder automation system.," *In 2020 IEEE Power and Energy Society General Meeting (PESGM)*, pp. 1–5, 2020.

[29] L. Zhaoxi and L. Wang, "Flipit game model-based defense strategy against cyber-attacks on scada systems considering insider assistance.," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 2791–2804, 2021.

[30] S. Hasan, A. Dubey, G. Karsai, and X. Koutsoukos, "A game-theoretic approach for power systems defense against dynamic cyber-attacks.," *International Journal of Electrical Power and Energy Systems*, vol. 115, 2020.

[31] W. Bi, C. Chen, and K. Zhang, "Optimal strategy of attack-defense interaction over load frequency control considering incomplete information," *IEEE Access*, vol. 7, pp. 75 342–75 349, 2019.

[32] S. Lakshminarayana and D. K. Yau, "Cost-benefit analysis of moving-target defense in power grids," *IEEE Transactions on Power Systems*, vol. 36, no. 2, pp. 1152–1163, 2021. DOI: 10.1109/TPWRS.2020.3010365.

[33] Y. Xiang and L. Wang, "A game-theoretic study of load redistribution attack and defense in power systems," *Electric Power Systems Research*, vol. 151, pp. 12–25, 2017.

[34] M. Esmalifalak, G. Shi, Z. Han, and L. Song, "Attack against electricity market-attacker and defender gaming.," *IEEE Global Communications Conference (GLOBECOM)*, pp. 3147–3152, 2012.

[35] D. Tang, Y. Fang, and E. Zio, "A zero-sum Markov defender-attacker game for modeling false pricing in smart grids and its solution by multi-agent reinforcement learning.," *29th European Safety and Reliability Conference (ESREL2019)*, 2019.

[36] Z. Zhang, S. Huang, Y. Chen, B. Li, and S. Mei, "Cyber-physical coordinated risk mitigation in smart grids based on attack-defense game.," *IEEE Transactions on Power System*, vol. 37(1), pp. 530–542, 2021.

[37] E. Karangelos and L. Wehenkel, "An iterative AC-SCOPF approach managing the contingency and corrective control failure uncertainties with a probabilistic guarantee," *IEEE Transactions on Power Systems*, vol. 34, no. 5, pp. 3780–3790, 2019. DOI: 10.1109/TPWRS.2019.2902486.

[38] B. Zeng and L. Zhao, "Solving two-stage robust optimization problems using a column-and-constraint generation method," *Operations Research Letters*, vol. 41, no. 5, pp. 457–461, 2013, ISSN: 0167-6377. DOI: https://doi.org/10.1016/j.orl.2013.05.003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167637713000618.

[39] W. Yuan, L. Zhao, and B. Zeng, "Optimal power grid protection through a defender–attacker–defender model," *Reliability Engineering & System Safety*, vol. 121, pp. 83–89, 2014.

[40] C. F. Camerer, T.-H. Ho, and J.-K. Chong, "A cognitive hierarchy model of games," *The Quarterly Journal of Economics*, vol. 119, no. 3, pp. 861–898, 2004.

[41] L. Che, X. Liu, Z. Li, and Y. Wen, "False data injection attacks induced sequential outages in power systems," *IEEE Transactions on Power Systems*, vol. 34, no. 2, pp. 1513–1523, 2018.

[42] M. Tian, M. Cui, Z. Dong, X. Wang, S. Yin, and L. Zhao, "Multilevel programming-based coordinated cyber physical attacks and countermeasures in smart grid," *IEEE Access*, vol. 7, pp. 9836–9847, 2019.

A

Load redistribution cyber-physical attack formulation

This appendix is based on publication [21] and is organized as follows. Section A.1 presents the mathematical notation that is used exclusively for the expressions shown in this appendix. The reader should kindly excuse the inevitable redefinition of the meaning of several symbols with respect to the main body of this report. Section A.2 verbally introduces the decision making problem of a malicious adversary seeking to provoke physical insecurity in the power grid by means of a load redistribution (measurement based) attack. Next, section A.3 presents the detailed mathematical formulation and establishes the correspondence with the abstract symbolic expressions used in the main body of this report. Finally, section A.4 presents a demonstrative implementation.

A.1 Notation

$\mathscr{G}$ set of generating units;

$\mathscr{L}$ set of transmission branches;

$\mathscr{N}$ set of nodes;

r_ℓ upper-level continuous variable, measuring the magnitude of the branch overloads induced by the attack;

$u_\ell^{\cdot}$ upper-level binary variable, indicating the overload status of branch ℓ, with superscripts $(+/-)$ for an overloaded branch in the positive/negative flow direction or 0 for no overload;

U parameter, modeling the minimum number of overloaded branches targeted by the attacker;

a_n upper-level binary variable, indicating the injection of false data at the load measurement of node n;

A parameter, modeling the attacker's available budget for attacking the grid load meters;

δd_n upper-level continuous variable, modeling the false active power demand measurement data injected by the attacker at node n;

ϵ parameter, modeling the maximum relative amount of false load measurement data that can be injected by the attacker;

d_n parameter, modeling the active power demand at node n;

$\gamma_{g,n}$ parameter, modeling the connectivity of generator g with node n;

p_{g0} parameter, modeling the dispatch of generator g;

p_g lower-level continuous variable, modeling the active power redispatch of generator g by the grid-operator;

$\lambda_{\ell,n}$ parameter, modeling the connectivity of branch ℓ with node n and the assumed flow direction;

f_ℓ^{ca} upper-level continuous variable, modeling the cyber-attacker's perceived active power flow value through branch ℓ;

X_ℓ parameter, modeling the reactance of branch ℓ;

θ_n^{ca} upper-level continuous variable, modeling the cyber-attacker's perceived voltage angle value at node n;

ρ_ℓ parameter, modeling the minimum threshold of overloaded flow per branch targeted by the attacker;

$\overline{f}_\ell$ parameter, modeling the capacity of branch ℓ;

M a large constant parameter;

c_g parameter, modeling the non-negative upward redispatch marginal cost of generator g;

π_g lower-level continuous variable, modeling the upward redispatch of generator g;

$\underline{p}_g$ parameter, modeling the minimum stable output of generator g;

$\overline{p}_g$ parameter, modeling the capacity of generator g;

f_ℓ^{go} lower-level continuous variable, modeling the grid-operator's perceived active power flow value through branch ℓ;

θ_n^{go} upper-level continuous variable, modeling the grid-operator's voltage angle value at node n.

A.2 Problem description

We model a malicious cyber-attacker seeking to maximize the grid physical insecurity through a load redistribution attack. More specifically, we consider an attacker falsifying bus load measurements so as to mislead the grid-operator into perceiving the grid as insecure and implementing unnecessarily generation redispatch actions. The cyber-attacker's objective is to maximize the total magnitude of branch overloads caused by the injection of false measurements and the resulting generation redispatching of the mislead grid-operator. Further, we consider that the cyber-attacker is seeking to induce an 'overwhelming' instance of power grid insecurity in terms of the minimum number and minimum magnitude of overloaded transmission branches.

A.3 Problem formulation

Objective function (A.1) seeks to maximize the total magnitude of the branch overloads induced by the cyber-attack. It is of course an instance of function (4.14) wherein the branch overload magnitude vector represents the power system state as perceived by the cyber-attacker.

Expression (A.2) imposes a limit on the maximum number of load meters that can be manipulated by the attacker given the access it has already established while using its manpower and computational resources. It is therefore an instance of constraint (4.15) from the compact formulation. Inequalities (A.3) enforces that the false load measurement data injection is balanced across the grid and (A.4) sets the maximum relative amount of false data that can be injected by the attacker at any node. As discussed in [41] constraints (A.3,A.4) are the standard *proxy* constraints for the undetectability of a load redistribution attack in the DC model. In other words, these describe the space of feasible attack vectors given the currently active detection countermeasures and correspond to an upper bound on the last term of (4.17) of the compact formulation.

Equality constraints (A.6) and (A.7) are the nodal power balance and branch flow definition formulas under the DC power flow approximations. These equalities correspond to the so-called cyber-attacker's physical model for the electric power grid, which was represented within (4.16) in the compact formulation. Notice that the power balance constraint (A.6) includes the optimal values of the generation redispatch variables ($p_g^\star$) as decided in the grid-operator's lower-level problem (A.20 – A.26). The group of inequalities (A.8 – A.12) is used to flag overloaded branches either in the positive or in the negative flow direction, while (A.13 – A.17) are used to measure the magnitude of the branch overloads caused by the cyber attack. Here we use parameter ($\rho_\ell \geq 1$) to model that a malicious cyber-attacker may strategically prefer to cause an overloaded flow larger than a threshold on every overloaded branch in order to create an overwhelming grid insecurity instance. We also introduce inequality constraint (A.19) to model that a malicious cyber-attacker may strategically prefer to overload at least a minimum number of branches ($U \geq 2$) in order to create an overwhelming grid insecurity instance outside the *"comfort zone"* of N-1 security. In fine, (A.6 – A.19) are a complete instance of the so-called severity constraints (4.16) of the compact problem formulation.

The lower-level problem (A.20 – A.26) is a standard DC-OPF problem modeling the reaction of the grid-operator to the injection of the false data by the attacker, seeking to minimize the cost of upward generation redispatching so as to maintain all perceived (*i.e.*, false) branch flow values within the respective capacity ratings. The cyber-attacker's decision strategy appears as the false data injection variable (e_n) in the right-hand-side of the power balance constraint (A.24), while supersrcipt (^{go}) denotes the (false) branch flow and voltage angle values perceived by the grid-operator.

We must finally underline that equality constraint (4.17) of the compact formulation, modeling the effect of the attack vector on the functionality of the cyber sub-system is directly integrated in the right-hand-side of power balance constraint (A.24). Indeed, as seen in this expression the load measurements available to the system operator are defined as the addition of the original values and the malicious data introduced by the cyber-attacker.

$$\max \sum_{\ell \in \mathscr{L}} r_\ell \tag{A.1}$$

$$\sum_{n \in \mathscr{N}} a_n \leq A \tag{A.2}$$

$$\sum_{n \in \mathscr{N}} e_n = 0 \tag{A.3}$$

for all nodes $n \in \mathscr{N}$:

$$-a_n \cdot \epsilon \cdot d_n \leq \delta d_n \leq a_n \cdot \epsilon \cdot d_n \tag{A.4}$$

$$a_n \in \{0, 1\} \tag{A.5}$$

$$\sum_{g \in \mathscr{G}} \gamma_{g,n} \left(p_{g0} + p_g^\star \right) - \sum_{\ell \in \mathscr{L}} \lambda_{\ell,n} \cdot f_\ell^{ca} = d_n \tag{A.6}$$

for all branches $\ell \in \mathscr{L}$:

$$f_\ell^{ca} = (1/X_\ell) \cdot \sum_{n \in \mathscr{N}} \lambda_{\ell,n} \cdot \theta_n^{ca} \tag{A.7}$$
$$u_\ell^+ + u_\ell^- + u_\ell^0 = 1 \tag{A.8}$$
$$f_\ell^{ca} - \rho_\ell \cdot \overline{f}_\ell \le u_\ell^+ \cdot M \tag{A.9}$$
$$f_\ell^{ca} - \rho_\ell \cdot \overline{f}_\ell \ge (u_\ell^+ - 1) \cdot M \tag{A.10}$$
$$-f_\ell^{ca} - \rho_\ell \cdot \overline{f}_\ell \le u_\ell^- \cdot M \tag{A.11}$$
$$f_\ell^{ca} + \rho_\ell \cdot \overline{f}_\ell \ge (1 - u_\ell^-) \cdot M \tag{A.12}$$
$$r_\ell \le (1 - u_\ell^0) \cdot M \tag{A.13}$$
$$(u_\ell^+ - 1) \cdot M + (f_\ell^{ca} - \overline{f}_\ell) \le r_\ell \tag{A.14}$$

$$r_\ell \le (1 - u_\ell^+) \cdot M + (f_\ell^{ca} - \overline{f}_\ell) \tag{A.15}$$
$$(u_\ell^- - 1) \cdot M - (f_\ell^{ca} + \overline{f}_\ell) \le r_\ell \tag{A.16}$$
$$r_\ell \le (1 - u_\ell^-) \cdot M - (f_\ell^{ca} + \overline{f}_\ell) \tag{A.17}$$
$$u_\ell^+, u_\ell^-, u_\ell^0 \in \{0, 1\} \tag{A.18}$$
$$\sum_{\ell \in \mathscr{L}} \left(u_\ell^+ + u_\ell^-\right) \ge U \tag{A.19}$$

subject to the model of the mislead grid-operator:

$$\min \sum_{g \in \mathscr{G}} c_g \cdot \pi_g \tag{A.20}$$

for all generators $g \in \mathscr{G}$:

$$\pi_g \ge 0 \tag{A.21}$$
$$\pi_g \ge p_g^\star \tag{A.22}$$
$$(\underline{p}_g - p_{g0}) \le p_g^\star \le (\overline{p}_g - p_{g0}) \tag{A.23}$$

for all nodes $n \in \mathscr{N}$:

$$\sum_{g \in \mathscr{G}} \gamma_{g,n} \left(p_{g0} + p_g^\star\right) - \sum_{\ell \in \mathscr{L}} \lambda_{\ell,n} f_\ell^{go} = d_n + \delta d_n \tag{A.24}$$

for all branches $\ell \in \mathscr{L}$:

$$f_\ell^{go} = (1/X_\ell) \cdot \sum_{n \in \mathscr{N}} \lambda_{\ell,n} \cdot \theta_n^{go} \tag{A.25}$$
$$-\overline{f}_\ell \le f_\ell^{go} \le \overline{f}_\ell. \tag{A.26}$$

A.4 Demonstrative implementation

A.4.1 Test case setup

We adopt the single-area version (24 bus) of the IEEE-RTS96 benchmark[1]. Following the practice of relevant studies (*e.g.*, [18], [19], [42]) we simulate a stressed operational condition by reducing

[1] All system data can be found at `https://matpower.org/docs/ref/matpower5.0/case24_ieee_rts.html`.

all branch transmission capacities to 65% of the original values. We further model a malicious cyber-attacker seeking to overload at least $U = 2$ transmission elements to at least $\rho_\ell = 5\%$ of the respective capacities. We set the cyber-attacker's resource constraint to falsifying at most $A = 10$ distinct load measurements and the maximum relative amount of false data per measurement to $\epsilon = 20\%$.

A.4.2 Perfect information load redistribution attack

Under the assumed conditions a cyber-attacker with perfect information, solving model (A.1 – A.26) with the correct values for all grid parameters, would indeed be able to induce 2 overloads in the grid by more than 5% of the respective branch capacities. More specifically, the cyber-attacker would provoke erroneous redispatch by the grid-operator eventually overloading branch 12 to 109.1% of its capacity and branch 23 to 118.6% of its capacity. The total magnitude of measurable overloads (*i.e.*, above the 5% threshold) would amount to 48.8 MW. Figure A.1 illustrates the optimal attack vector, with the x-axis showing the index of the affected bus load meter and the y-axis the percentage of change in the falsified load data.

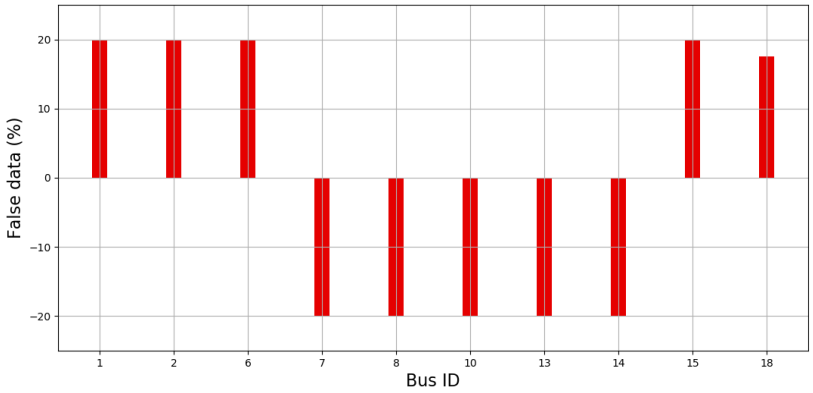

Figure A.1: Perfect information optimal attack vector

A.4.3 Cyber-attacks with imperfect information on the grid admittances only

We start by considering that realistic cyber-attackers may rely on inaccurate data considering the grid admittances only. To do so, we derive 10000 inaccurate grid samples, by applying a distinct error term to the admittance value of each branch, which is uniformly distributed in the range $\pm 10\%$. Performing the respective simulations, we found that such (moderate) inaccuracy translates into 2677 (out of 10000) unique load redistribution attack vectors, with an average impact (*i.e.*, total measurable overload) of 28.36 MW. The histogram in Fig. A.2 shows the distribution of the impact of such potential attacks, which as anticipated ranges from 0 (for the case of not attempted or failed attacks) to the upper-bound set by the perfect information cyber-attack.

We further assess the risk through the pie-chart in Fig. A.3. As shown in this chart, due to the assumed informational imperfections a cyber-attacker would only be able to correctly identify the optimal perfect information attack vector from Fig. A.1 on 23.4% of the simulated instances. Conversely, on 15% of the sampled instances a cyber-attacker would falsely believe that it would be fruitless to launch any load redistribution attack while on 6.5% of the instances, she would launch an attack that would not be harmful to the grid. Observing that on 40.9% of the instances a cyber-attack with imperfect information would cause an overflow on at least two grid branches, while on 78.5% it would cause an overflow on at least one branch, we may infer from Fig. A.3 that this system is insecure[2].

[2] One may notice however that informational imperfections are in favor of security, as a perfectly informed attacker would be able to induce insecurity with 100% likelihood.

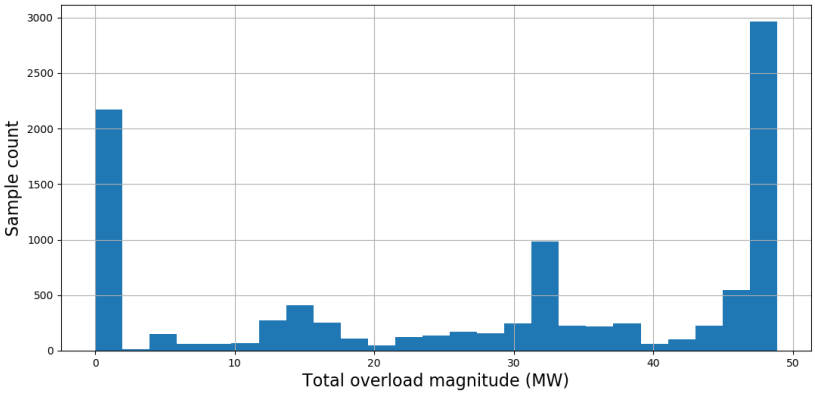

Figure A.2: Impact distribution of cyber-attacks with imperfect admittance data

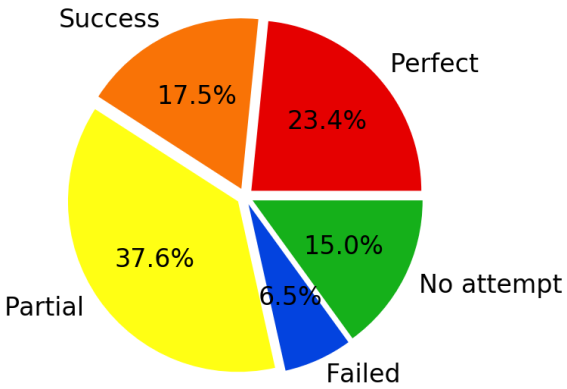

Figure A.3: Classification of cyber-attacks with imperfect admittance data

Pursuing the analysis from a risk management perspective, Fig. A.4.a shows the relative frequency of attacking each distinct bus load meter amongst the 10000 sampled cyber-attacks. The bars in blue correspond to the perfect information optimal attack vector from Fig. A.1 and it is notable that these are the meters ranked first in order of decreasing frequency. Specifically, the least-frequently attacked meter from those in the perfect information optimal vector has been selected in 58% of imperfect attacks while the most-frequently attacked meter from those not in the perfect information optimal vector has only been selected in 41% of the imperfect attacks. Further, as illustrated further in Fig. A.4.b, 97.5% of the imperfect cyber-attacks share at least 7 common attacked asset(s) with the perfect information cyber attack while all 10 meters from Fig. A.1 have been attacked in 39.5% of the sampled instances. The important take-away here is that protecting the meters that would have been attacked in the perfect information case may well be sufficient to detect and prevent with very high probability the cyber-attacks under imperfect information from physically harming the system.

Finally, Fig.A.5 demonstrates which transmission branches would be overloaded due to the imperfect cyber-attacks. Adopting the color-coding of Fig. A.3, we show that for a large share of the samples the imperfect cyber-attack results in overloading the same branches as the perfect information attack, albeit to a smaller degree. The take-away here is that taking physical preventive/corrective measures for the possible joint outage of these branches could also be an effective strategy for managing cyber-physical risk. Notice the small frequency of imperfect cyber-attacks overloading three branches, which are suboptimal in terms of total overload magnitude.

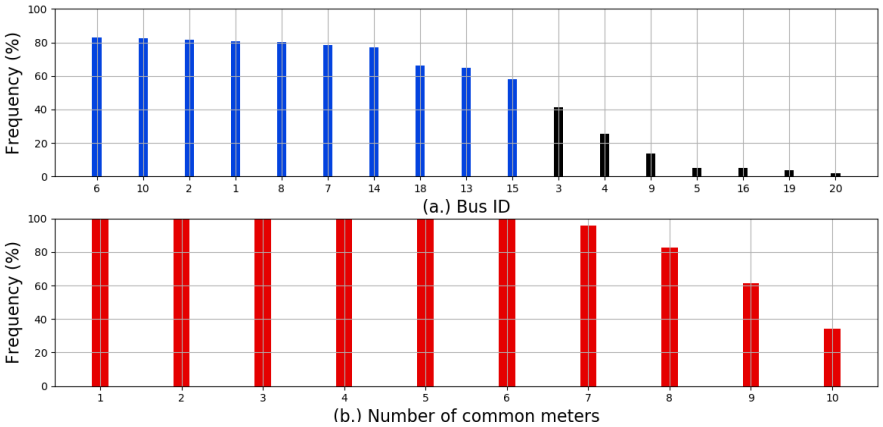

Figure A.4: Frequency of attacks (a.) per meter and (b.) sharing common meters with the perfect information attack

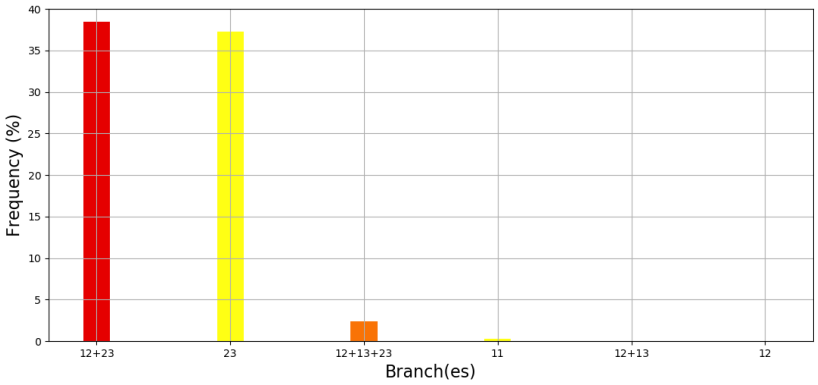

Figure A.5: Physical impact of of cyber-attacks with imperfect admittance data

A.4.4 Sensitivity analysis with respect to the admittance error range

To validate the aforementioned observations we perform a sensitivity analysis by drawing two additional samples of 10000 inaccurate grid instances while assuming that the imperfect cyber-attacker's error in admittance values is uniformly distributed in the $\pm 5\%$ and $\pm 15\%$ ranges. As anticipated, in the former case the average impact of the imperfect cyber-attacks increases to 35.6 MW (with 1428 unique attack vectors) while in the latter it reduces slightly to 26.72 MW (with 4044 unique attack vectors). It is noteworthy that in the case of reduced inaccuracy, Fig. A.6.a., the percentage of so-called *perfect* attacks more-than doubles to 51.3%. This shows that (the reduced) inaccuracy has a smaller effect on the attack vector of the imperfect cyber-attacker. Conversely, in Fig. A.6.b., increased inaccuracy almost halves the percentage of *perfect attacks*, with the most notable increase observed in the *partial* attack class.

In our detailed results we further find that for both cases (*i.e.*, under reduced or increased randomness) the set of meters included in the optimal perfect information attack vector from Fig. A.1 remains the set of the 10 most frequently attacked meters, while the frequency of attacking a large subset of these meters remains as high. Specifically, for the $\pm 5\%$ error range 99.7% of the imperfect attacks share at least 7 meters in common with the perfect information attack and for the $\pm 15\%$ error range this percentage only reduces to 95.4%. These findings are well in line with the argument that protecting the meters involved in the perfect information cyber-attack is a good starting point for detecting and preventing any random imperfect cyber-attack vector. Similarly, concerning the branches

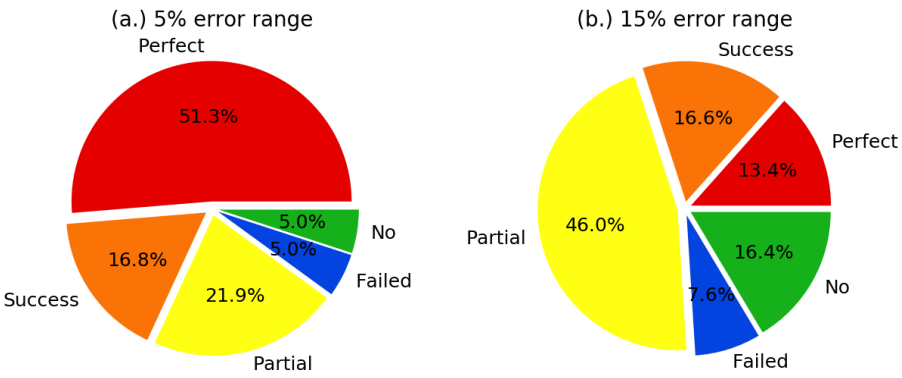

Figure A.6: Classifications for (a.) $\pm 5\%$ and (b.) $\pm 15\%$ admittance error

that may undergo overloads in the aftermath of an imperfect cyber-attack, our sensitivity analysis detailed results qualitatively follow the representation of Fig. A.5. That is, most frequently both branches that would be overloaded in the case of the perfect cyber-attack are also affected by the imperfect cyber-attacks.

A.4.5 Cyber-attacks with imperfect information on the branch capacities only

We continue the analysis by henceforth considering the case where the cyber-attacker relies on inaccurate data about the branch capacities only. We sample additionally 10000 inaccurate grids, by applying a distinct error term to the capacity value of each branch, which is again uniformly distributed in the range $\pm 10\%$. With such assumptions, the average cyber-attack impact reduces to 25.31 MW while the number of unique cyber-attacks increases to 6737.

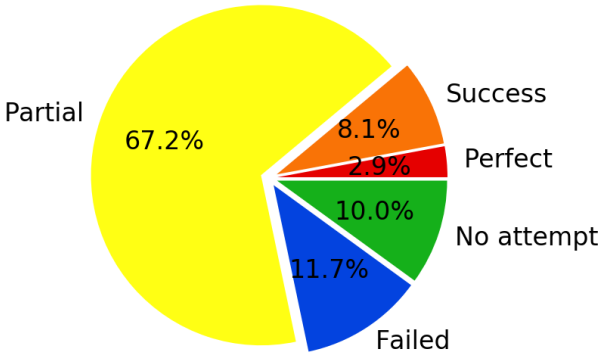

Figure A.7: Classification of cyber-attacks with imperfect capacity data

Fig. A.7 presents the classification of the random cyber-attacks. The qualitative difference with respect to imperfect admittance values is striking in comparison to Fig. A.3. Indeed, for the same error range: i) the share of *perfect* attacks has collapsed, ii) the share of *success* attacks is (more than) halved, iii) the share of *partial* attacks is considerably increased, and, iv) the share of ineffective attacks is moderately increased. In other words, imperfect information on the branch capacities leads to much less effective cyber-attacks posing a smaller risk to the system cyber-physical security.

We can identify systematic reasons for this finding. Indeed, in case the cyber-attacker undervalues branch capacities, she is prone to overestimating the impact of an attack vector in firstly misleading

the grid-operator to redispatch generation to avoid overloads under the load redistribution, and secondly in causing actual overloads by way of the erroneous redispatch. This explains the large shift from *perfect/success* to *partial* attacks. Also, in case the cyber-attacker overvalues branch capacities, she is prone to believing there is no potential for attacking the grid.

Concerning risk management, we once again find that the frequency of attacking a large subset of meters identified in the perfect information attack remains indicative, with 97.8% of the imperfect attacks targeting at least 6 meters from the perfect information optimal vector and 87.3% of the imperfect attacks targeting at least 7 of these meters. As should be anticipated by the dominance of the *partial* attack category, the most frequent overflow in the system now concerns a single transmission branch, Fig. A.8. Notice here that the groups of affected branches (x-axis) are all in common with Fig. A.5. Both these findings further showcase the relevance of these groups of cyber and physical sub-system assets for preventive and corrective cyber-physical risk management.

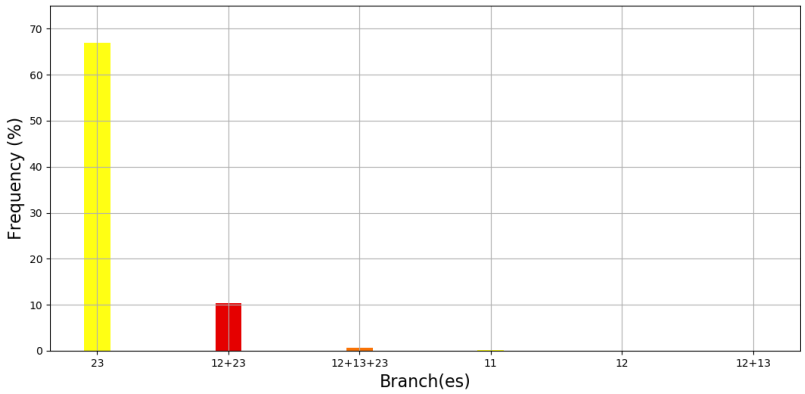

Figure A.8: Physical impact of of cyber-attacks with imperfect capacity data

This project is supported by the Belgian Energy Transition Fund