Cyber-physical risk modeling with imperfect cyber-attackers

Efthymios Karangelos and Louis Wehenkel,

Institut Montefiore,
Department of Electrical Engineering and Computer Science,
Universitè de Liège,
Liège, Belgium.

ETH PSL Seminar
08 December 2021

CYPRESS

With the support of
the Energy Transition Fund

economie
FPS Economy, S.M.E.s, Self-employed and Energy

haulogy
The Energy IT Company

KU LEUVEN

ULB UNIVERSITÉ LIBRE DE BRUXELLES

LIÈGE université
project coordination

Why CYPRESS?

- The modern EPS is a cyber-physical system:
  - SCADA/EMS, telecommunications, "smart-grid" solutions on the system side;
  - smart-homes, distributed generation on the end-user side.
- In addition to physical threats (e.g., contingencies) it is under risk from ...
  - the cyber vulnerabilities (e.g., software bugs),
  - malicious cyber-attackers seeking to disrupt the supply of electricity.
- Going from physical to cyber-physical risk management requires ...
  - (co-)simulating the cyber and physical sub-systems;
  - modeling the strategies of all involved actors, including malicious cyber-attackers!

Malicious cyber-attacker modeling (cf., [1])

Planner aims to prevent the success of potential attacks.

Attacker aims to disrupt the system functionality.

Operator aims to maintain the system functionality after any attack.

- Deterministic max min optimization:

  max a (perfect) attacker, fully aware of the properties of the system and of its operator;

  min an operator optimally responding to the sustained cyber-attack.

⋆ Solving these deterministic bi-level problems not trivial for realistic systems!

E.g.: Load redistribution (false data injection) modeling

max attacker tampers with the load data received by the control center;

- subject to resource & attack undetectability constraints;
- and to the operator's decision making model.

E.g.: Load redistribution (false data injection) modeling

max attacker tampers with the load data received by the control center;

- subject to resource & attack undetectability constraints;
- and to the operator's decision making model.

min operator reacts to the perceived system state by redispatching generation;

- based on false load data;
- subject to the power flow model & the system constraints.

E.g.: Load redistribution (false data injection) modeling

max attacker tampers with the load data received by the control center;

- subject to resource & attack undetectability constraints;
- and to the operator's decision making model.

min operator reacts to the perceived system state by redispatching generation;

- based on false load data;
- subject to the power flow model & the system constraints.

- The system ends-up being operated:
  - uneconomically, if generation is redispatched out of merit,
  - or even insecurely, if the actual system state violates its limits.

Realistic cyber-attackers have imperfect information

- Realistic attacks will be based on (randomly) inaccurate grid data [2, 3];

  e.g. a realistic attacker can't observe and react instantaneously to the status of every circuit breaker, tap-changer, etc..

Realistic cyber-attackers have imperfect information

- Realistic attacks will be based on (randomly) inaccurate grid data [2, 3];

  e.g. a realistic attacker can't observe and react instantaneously to the status of every circuit breaker, tap-changer, etc..

- Is this relevant for cyber-physical risk assessment?
  - should we study a distribution of random attackers rather than the perfect information worst case?

- Is this relevant for cyber-physical risk management?
  - should we state stochastic rather than deterministic min max min problems?

In this work . . .

- We propose a new formulation for load-redistribution cyber-physical attacks:
  - seeking to maximize the magnitude of branch overloads;
  - while ensuring that the grid security will be severely compromised.
- We analyze the distribution of attacks designed with randomly inaccurate data:
  - discussing implications for risk assessment and risk control.

1. The cyber-attack optimization problem formulation.

---

2. Modeling imperfect information cyber-attacks.
3. Results & discussion.

The max min objectives

max Attacker's objective is the total magnitude of branch overloads induced by:

- the false load demand measurements;
- and the corresponding generation redispatch by the (mislead) grid operator.

---

min Operator's objective is the cost of generation redispatching:

  - given the false load data,
  - so as to keep the perceived (fake) system state within limits.

The complete formulation is available as an appendix to these slides, and at `https://arxiv.org/abs/2110.00301`.

The attacker's constraints

- Attack undetectability (linear):
  - net false data injection is balanced across the system;
  - false data injection per grid bus is bounded.
- Attack resources (mix-integer linear):
  - total number of false measurements (attacked load buses) is upper bounded.

The attacker's constraints

- Attack undetectability (linear):
  - net false data injection is balanced across the system;
  - false data injection per grid bus is bounded.
- Attack resources (mix-integer linear):
  - total number of false measurements (attacked load buses) is upper bounded.
- Attack severity – new (mix-integer linear):
  - a lower bound on the number of branch overloads to be achieved;

    e.g. at least 2 branches;
  - a lower bound per branch on the measurable overload magnitude.

    e.g. overloading a branch at 100.0001 % is pointless.

The grid modeling constraints

- Attack physical-impact (mixed-integer linear):
  - nodal injections computed with the true load data & the operator's generation redispatching variables;
  - power balance, DC power flow;
  - generation redispatching variables optimally solve the operator's decision making problem;
    - given the false load data;
    - subject to power balance, DC power flow, branch capacity and generation capacity constraints;
    - reformulated through the KKT optimality conditions.

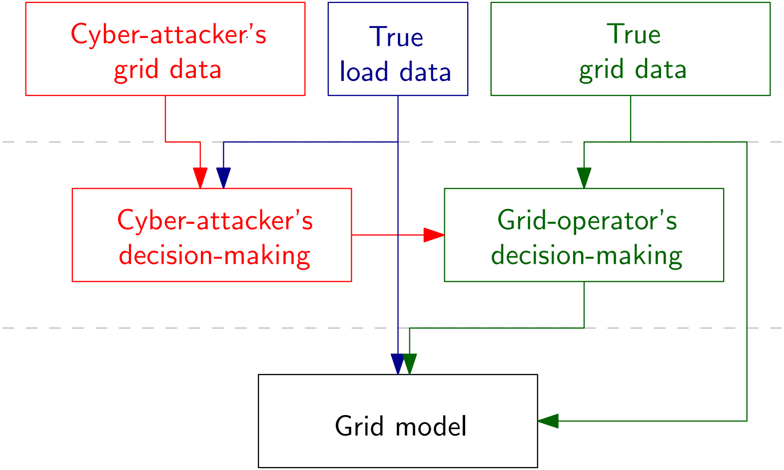
2. Modeling imperfect information cyber-attacks

3. Results & discussion.

Randomizing the cyber-attacker's grid data

- The attacker may be misinformed about ...
  - the branch admittances (depending on FACTs, PSTs, etc.);
  - the branch ratings (depending on ambient conditions, operator risk aversion etc).
- How do we model this?
  - applying a uniformly distributed error term on each distinctive data point;
  - assuming everything is equiprobable and sampling ahead.

Modeling flowchart



Evaluation sequence

- Given a (random) inaccurate grid data instance.
- Attacker & operator solve different decision-making problems.
  a. attacker uses the inaccurate grid data to define its attack vector;
  b. operator faces the attack (false load data) but uses the correct grid data to select its reaction.
- The system state needs to be recomputed with:
  - the operator's redispatching;
  - the actual load values;
  - the correct grid data.

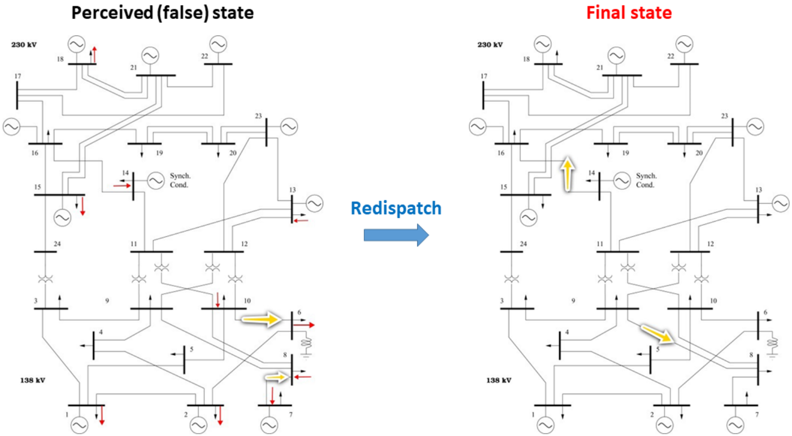
3. Results & discussion

Test case setup

- The single-area IEEE RTS 24;
  - branch ratings reduced to 65% to model system stress (common in this literature);
- The attacker's parameters;
  - can alter at most 10 load measurements;
  - can falsify any measurement with $\pm 20\%$ at most;
  - targeting at least 2 overloaded branches;
  - with at least 5% overload.

Benchmarking: the perfect information attack

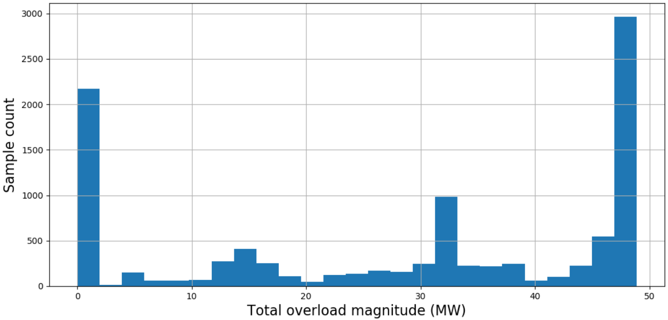
Benchmarking: the perfect information attack



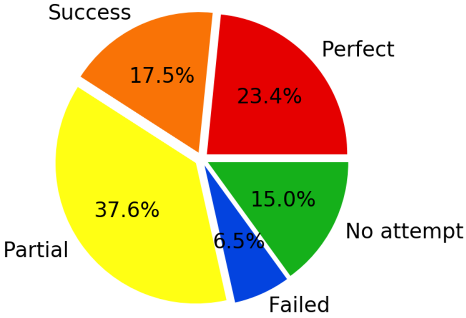
- Total overload magnitude is 48.8 MW.

Cyber-attacks with imperfect admittance data ($\pm 10\%$) only

- 2677 unique attacks out of 10000 samples;
- Average total overload magnitude is 28.36 MW ($\sim$ 58%).



Cyber-attacks with imperfect admittance data ($\pm 10\%$) only



Perfect information.

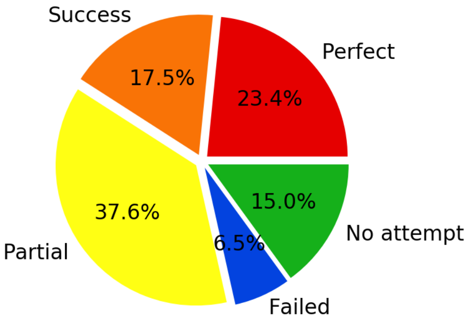
Success: only meet severity target.

Partial success: other physical impact.

Failure: no physical impact.

No attempt: perceived infeasible.

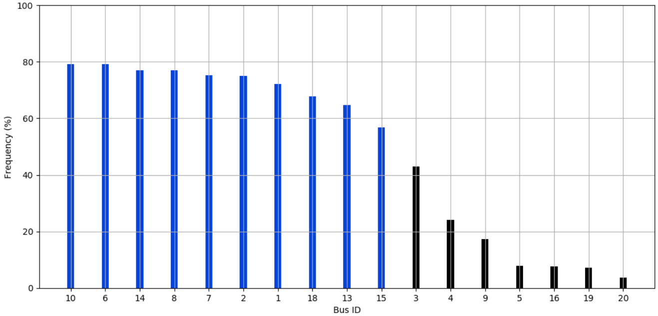
Cyber-attacks with imperfect admittance data ($\pm 10\%$) only



- Imperfections harm the cyber-attack;
  - only 40% of the imperfect info attacks meet the targeted severity.

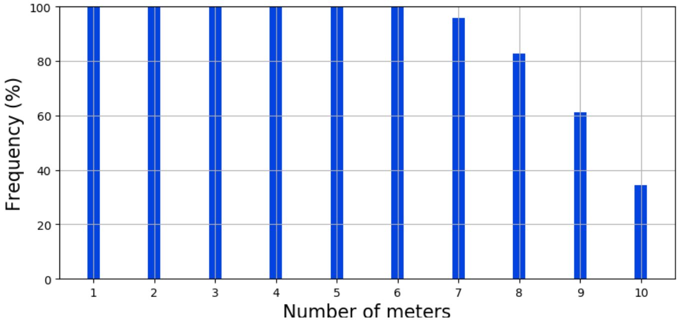
- The system looks insecure;
  - 78.5% of the imperfect info attacks have a physical impact.

Cyber-attacks with imperfect admittance data ($\pm 10\%$) only



- The buses targeted in the perfect information attack are most frequently attacked.

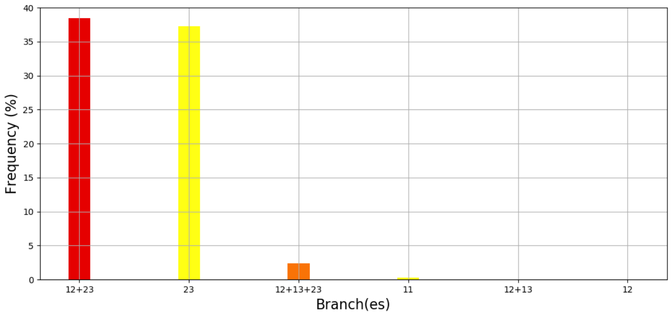
Cyber-attacks with imperfect admittance data ($\pm 10\%$) only



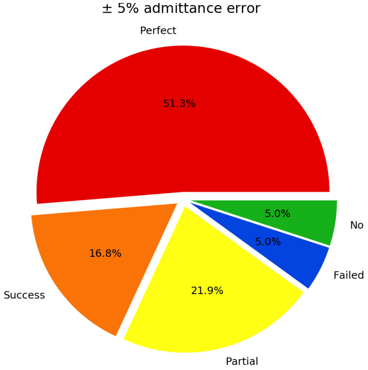
- All 10 buses selected in 39.2% of the attacks, at least one of these in all attacks.

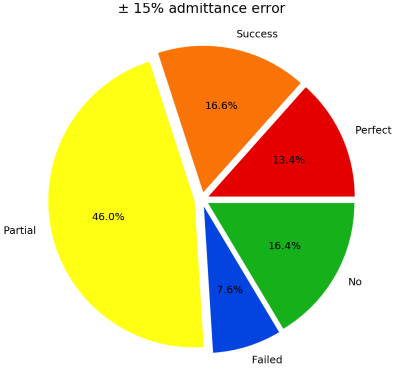
Cyber-attacks with imperfect admittance data ($\pm 10\%$) only

- The physical impact of these attacks is also coinciding.



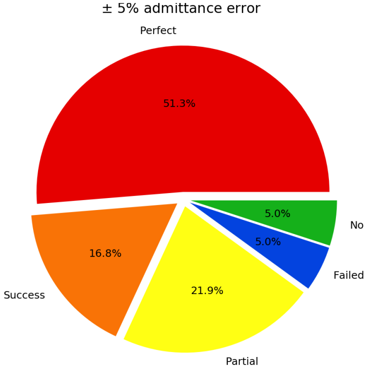
Cyber-attacks with imperfect admittance data only – sensitivity

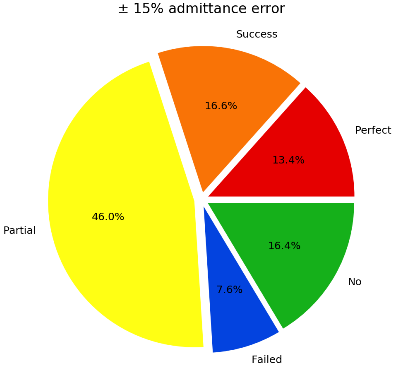




- Inaccuracy affects the potential to identify the perfect information attack;

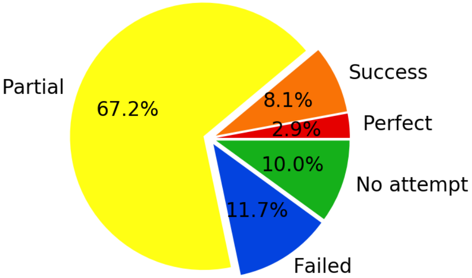
Cyber-attacks with imperfect admittance data only – sensitivity





- Inaccuracy affects the potential to identify the perfect information attack;
- but, no major change in terms of the buses targeted under the various attacks.

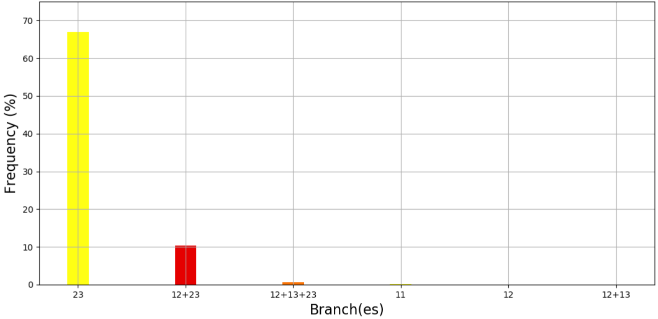
Cyber-attacks with imperfect **branch rating** data ($\pm 10\%$) only



- Much less effective attacks;
  - share of perfect attacks collapses;
  - share of partial attacks increases;
  - most attacks don't meet the attacker's standards.

- The system still looks insecure;
  - 78.2% of the imperfect info attacks have a physical impact.

Cyber-attacks with imperfect branch rating data ($\pm 10\%$) only



- The physical impact of these attacks is still coinciding;
- Affected branches (x-axis) is the same as in the case of inaccurate admittances.

The test-case take-aways

- Cyber-physical risk-assessment;
  - imperfect information wouldn't stop the cyber-attacker for physically disrupting the grid;
  - in spite of imperfections, the entry-points in the cyber-system are consistent with the perfect attack;
  - and the exit-points in the physical-system are also coincidental.
- Cyber-physical risk-management;
  - perfect information attack reveals effective priorities for preventive/corrective risk mitigation on the cyber and physical sub-systems.

Further work

- Generalizing over alternative test-systems;
  - Consistency in cyber/physical entry/exit points?
- Modeling alternative types of cyber-attackers;
  - different attack types and/or attack objectives;
  - stochastic bilevel optimization?
- From risk modeling to risk management;
  - $\min \max \min$ planner-attacker-operator under information uncertainty?

Thank you for your attention!

e.karangelos@uliege.be

References

[1] 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.

[2] 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.

[3] 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.

Implementation overview

- Decision-making models:
  - a MILP reformulation of the cyber-attacker *vs* operator $\max\min$ problem (using big-M for disjunctive inequalities);
  - an LP corresponding to the inner min for the operator's redispatching (DC-OPF).
- Grid model is a DC power flow.
- Developed in Julia/JuMP using the PowerModels.jl framework the CPLEX solver.

Problem formulation (1/4): attack properties

$$\max \sum_{\ell \in \mathcal{L}} r_\ell \tag{1}$$

$$\sum_{\ell \in \mathcal{L}} \left(u_\ell^+ + u_\ell^-\right) \geq U \tag{2}$$

$$\sum_{n \in \mathcal{N}} a_n \leq A \tag{3}$$

$$\sum_{n \in \mathcal{N}} e_n = 0 \tag{4}$$

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

$$-a_n \cdot \epsilon \cdot d_n \leq e_n \leq a_n \cdot \epsilon \cdot d_n \tag{5}$$

$$a_n \in \{0, 1\} \tag{6}$$

Problem formulation (2/4): true grid state

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

$$\sum_{g \in \mathcal{G}} \gamma_{g,n} \left(p_{g0} + p_g^\star \right) - \sum_{\ell \in \mathcal{L}} \lambda_{\ell,n} \cdot f_\ell^t = d_n \tag{7}$$

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

$$f_\ell^t = (1/X_\ell) \cdot \sum_{n \in \mathcal{N}} \lambda_{\ell,n} \cdot \theta_n^t \tag{8}$$

Problem formulation (3/4): overload sense & magnitude

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

$$u_\ell^+ + u_\ell^- + u_\ell^0 \leq 1 \tag{9}$$

$$f_\ell^t - \rho_\ell \cdot \overline{f}_\ell \leq u_\ell^+ \cdot M \tag{10}$$

$$f_\ell^t - \rho_\ell \cdot \overline{f}_\ell \geq (u_\ell^+ - 1) \cdot M \tag{11}$$

$$-f_\ell^t - \rho_\ell \cdot \overline{f}_\ell \leq u_\ell^- \cdot M \tag{12}$$

$$f_\ell^t + \rho_\ell \cdot \overline{f}_\ell \geq (1 - u_\ell^-) \cdot M \tag{13}$$

$$r_\ell \leq (1 - u_\ell^0) \cdot M \tag{14}$$

$$(u_\ell^+ - 1) \cdot M + (f_\ell^t - \overline{f}_\ell) \leq r_\ell \tag{15}$$

$$r_\ell \leq (1 - u_\ell^+) \cdot M + (f_\ell^t - \overline{f}_\ell) \tag{16}$$

$$(u_\ell^- - 1) \cdot M - (f_\ell^t + \overline{f}_\ell) \leq r_\ell \tag{17}$$

$$r_\ell \leq (1 - u_\ell^-) \cdot M - (f_\ell^t + \overline{f}_\ell) \tag{18}$$

$$u_\ell^+, u_\ell^-, u_\ell^0 \in \{0, 1\} \tag{19}$$

Problem formulation (4/4): mislead grid operator

$$p_g^\star \in \arg\min \sum_{g \in \mathcal{G}} c_g \cdot \pi_g \tag{20}$$

for all generators $\ell \in \mathcal{G}$:

$$0 \leq \pi_g \geq p_g \tag{21}$$

$$(\underline{p}_g - p_{g0}) \leq p_g \leq (\overline{p}_g - p_{g0}) \tag{22}$$

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

$$\sum_{g \in \mathcal{G}} \gamma_{g,n} (p_{g0} + p_g) - \sum_{\ell \in \mathcal{L}} \lambda_{\ell,n} f_\ell^f = d_n + e_n \tag{23}$$

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

$$f_\ell^f = (1/X_\ell) \cdot \sum_{n \in \mathcal{N}} \lambda_{\ell,n} \cdot \theta_n^f \tag{24}$$

$$-\overline{f}_\ell \leq f_\ell^f \leq \overline{f}_\ell. \tag{25}$$