SAND2019-3300PE

LDRD
Laboratory Directed Research and Development

Rigorous processes for cybersecurity experimentation: Sandia's SECURE project

Thomas D. Tarman

Sandia National Laboratories

tdtarma@sandia.gov

Supported by the Laboratory Directed Research and Development program at Sandia National Laboratories, a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525.

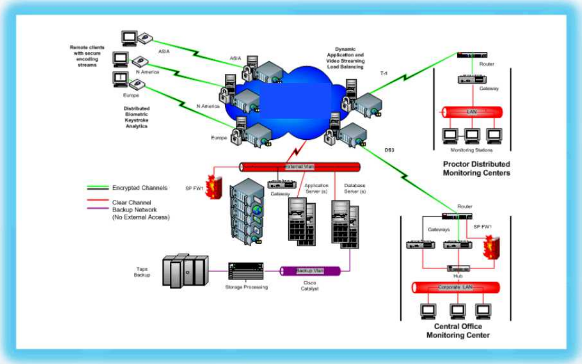
Outline

- Cyber experimentation background and tools
- An example and questions for this research
- Research overview - thrusts and questions
- Early results
- Conclusions

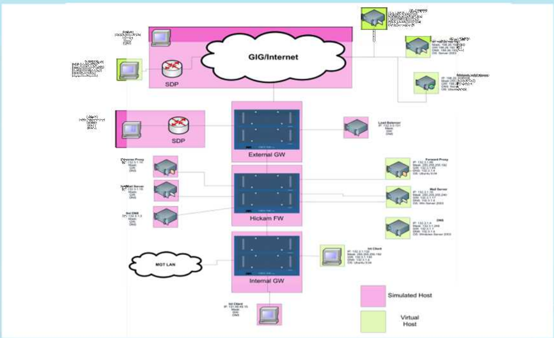
SECURE: Science and Engineering of Cybersecurity through Uncertainty Quantification and Rigorous Experimentation

Emulytics: Sandia’s tool suite for cybersecurity experimentation using emulation testbeds.

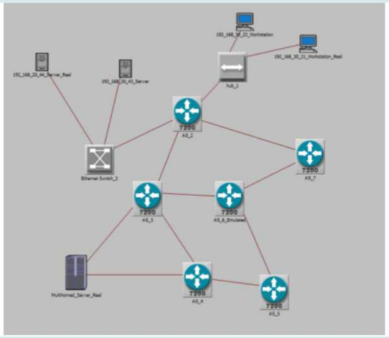
Cyber experimentation approaches



ACTUAL SYSTEM



VIRTUALIZED TESTBED

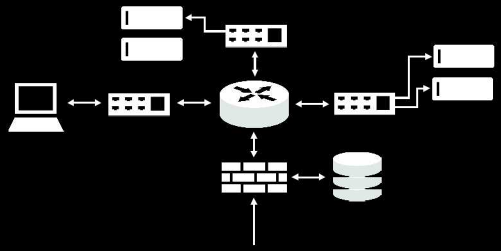


SIMULATION TESTBED

Interoperability in a single experiment

LIVE ← Increase Realism | Decrease Cost, Decrease Time → SIMULATED

REAL HARDWARE REAL SOFTWARE	ABSTRACT HARDWARE REAL SOFTWARE	ABSTRACT HARDWARE ABSTRACT SOFTWARE



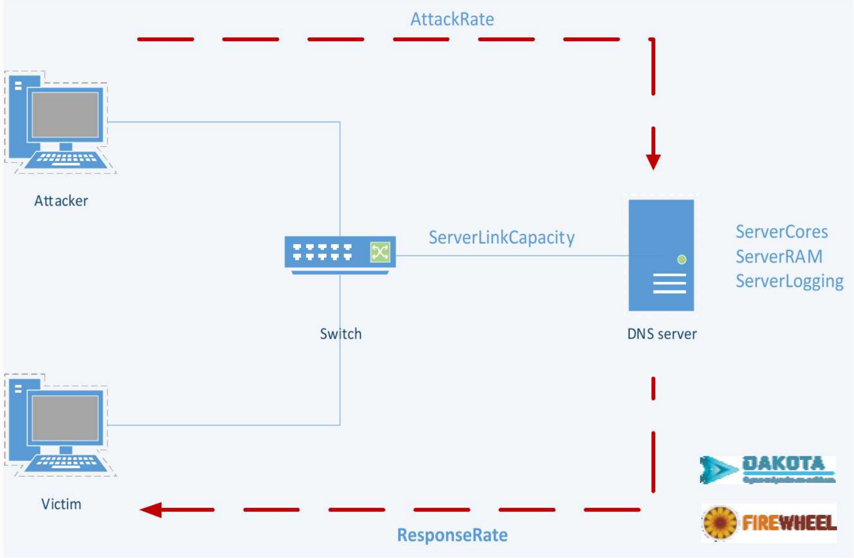
RTU – Remote Terminal Unit
PLC – Programmable Logic Controller

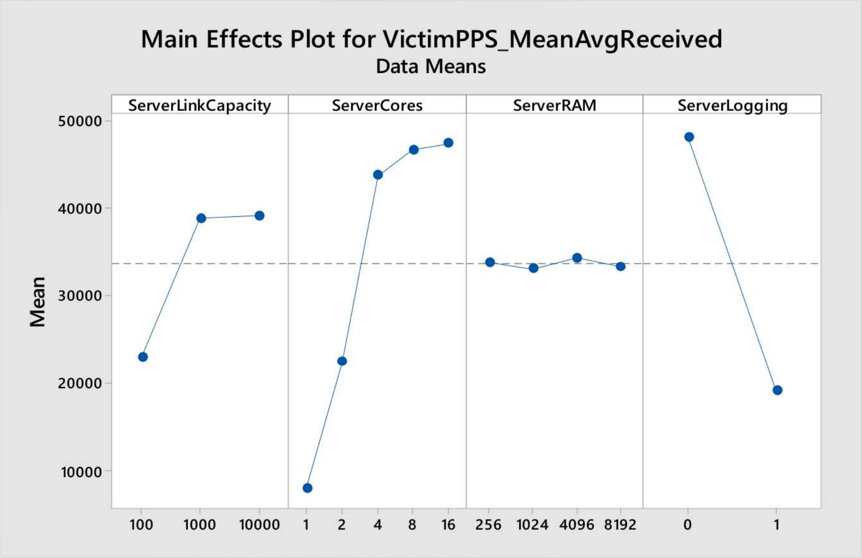
OPC – OLE for Process Control
SCADA – Supervisory Control And Data Acquisition

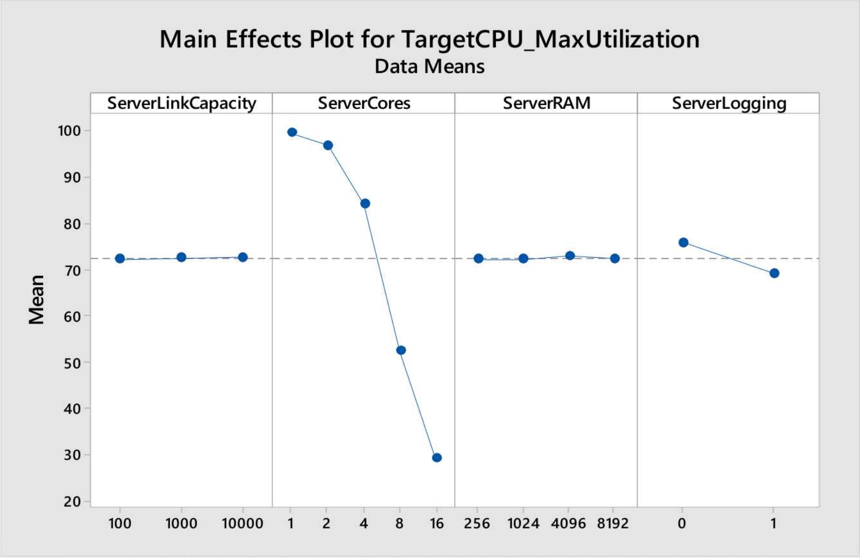
ICS – Industrial Control System

A simple, specific example – DNS amplification attack

- Threat – DNS request intensity (uncertain variable)
- Response metrics
  - Server CPU utilization
  - Amplified traffic to victim
- Questions
  - Sensitivity of outputs on inputs?
  - Parameters that optimize both responses?
  - Effect of threat uncertainty on results?







A simple, specific example – DNS amplification attack

Project questions:

- How does this scale? Can it be computationally efficient?
- Is the process generalizable? As complexity increases, can we sample effectively and build robust statistical models?

Emulytics team questions:

- Are there engineering hurdles associated with automated design of experiments and computational efficiency at scale?
- What are the research hurdles associated with modeling sophisticated (and often unknown) threats with uncertainty?
- How do we confidently make a V&V case at scale?

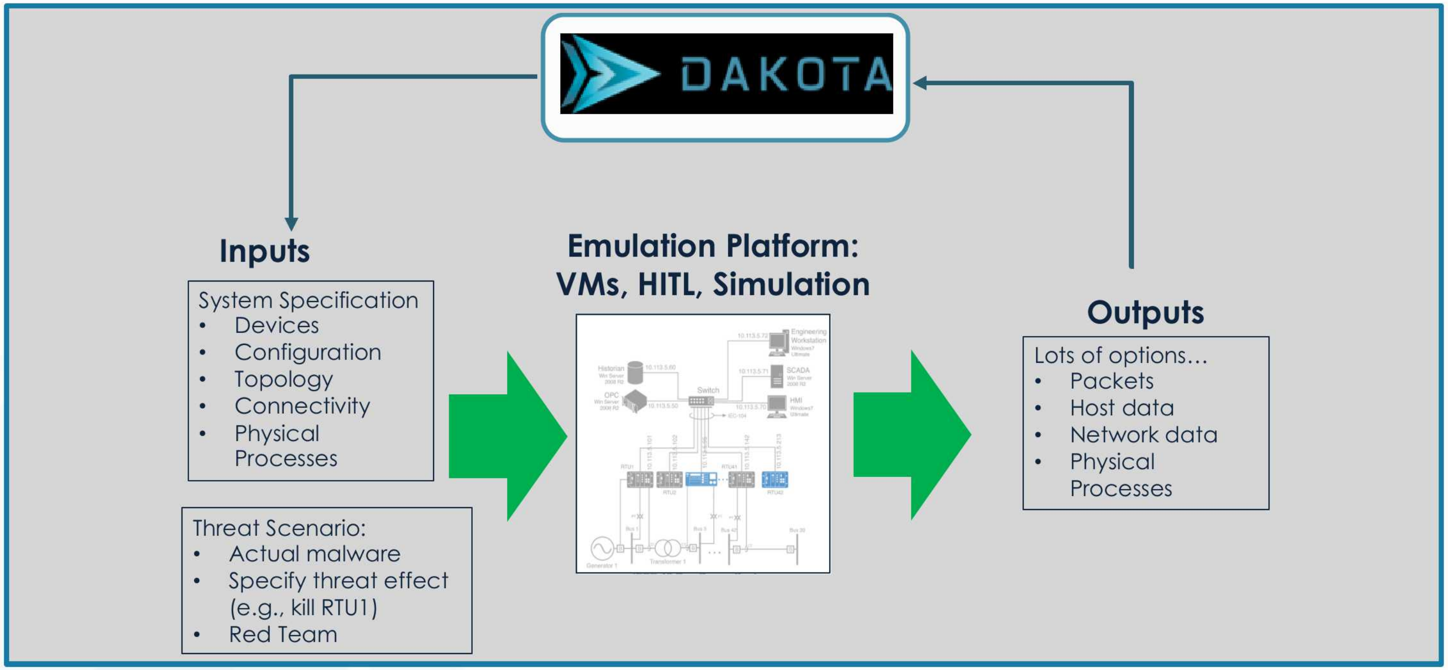
What does success look like?

Develop theory and tools (SECUREtk) that guide the cyber experimentalist to properly design, efficiently conduct, and analyze rigorous experiments, producing high quality data suitable for decisions about high-consequence systems.

- **STEPS**

1. Enhance emulation-based modeling processes and platforms
2. Develop methods for modeling uncertain threats
3. Quantitatively assess model confidence

Research thrust: Emulytics platform/modeling enhancements

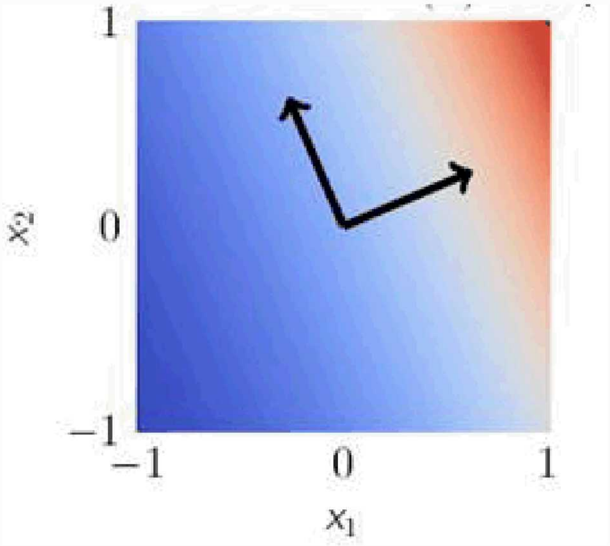


Question: Are there engineering hurdles associated with automated design of experiments and computational efficiency at scale?

- Develop exemplar and "toy model" questions and model
  - Toy model – Emulation model only (no PowerWorld component), to look at more complex (but relevant) cyber topologies
- Interfaces to allow external control over parameters and execution
- Experiment pause/resume/restart
  - Assess whether existing mechanisms are sufficient

Efficiency Improvements for UQ

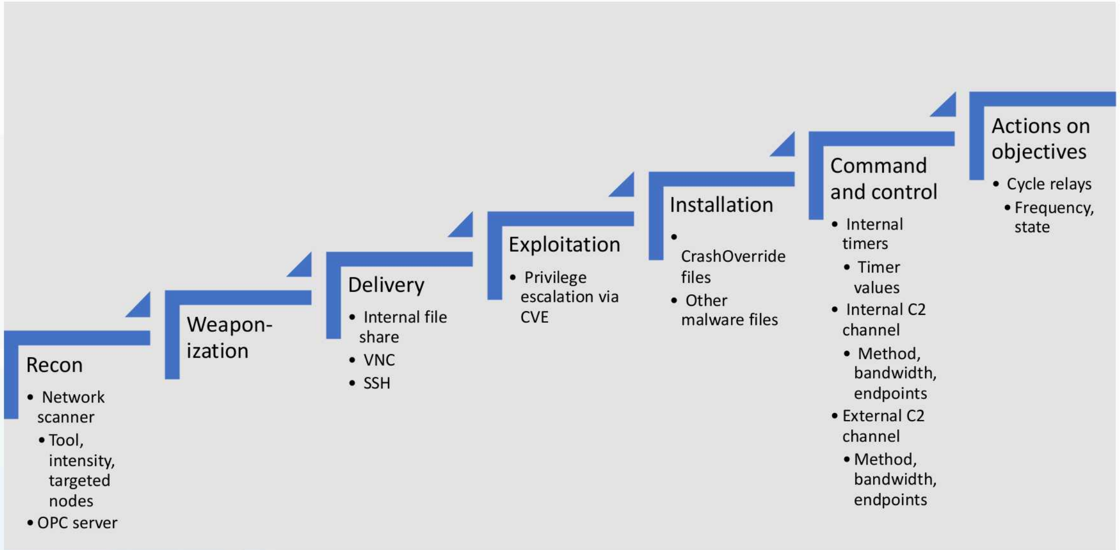
- **Dimension Reduction**
  - Determine a reduced or compressed representation of the Emulytic model's inputs and/or outputs.
  - Reduced space techniques involve a linear or nonlinear mapping between the full space to a reduced space of meta variables. Example: Principal components analysis (XPCA), active subspace



- **Multifidelity approaches**
  - Take a large number of low fidelity runs and a small number of high fidelity runs to achieve statistics on high fidelity responses
  - Relies on variance reduction: must have correlation between the low and high fidelity model
  - Active work on continuous problems→ translate to discrete

XPCA: eXtending PCA for Combinations of Discrete and Continuous Data, Kincher-Winoto, Kolda, and Anderson-Bergman, SAND2018-8213C. Also at: arXiv:1808.07510

Research thrust: Modeling uncertain threats



Adapted from: Hutchins, Eric, Michael Cloppert, and Rohan Amin. "Intelligence-Driven Computer Network Defense Informed by Analysis of Adversary Campaigns and Intrusion Kill Chains." *The Proceedings of the 6th International Conference on Information Warfare and Security*. 2011.

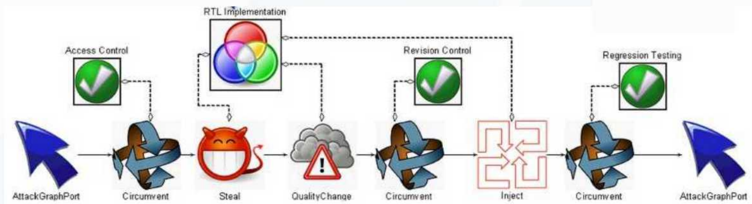
Question: What are the research hurdles associated with modeling sophisticated (and often unknown) threats with uncertainty?

- Specific threats evolve, so adopt frameworks that can be updated as threats change
  - Lockheed Martin Cyber Kill Chain
  - Graph-based Probabilistic Learning Attacker and Dynamic Defender (GPLADD)
  - Extensible threat modeling tools for emulation-based cyber experimentation
- Use GPLADD within CKC framework to inform threat/defense distributions and narrow parameter space for emulation-based experiments

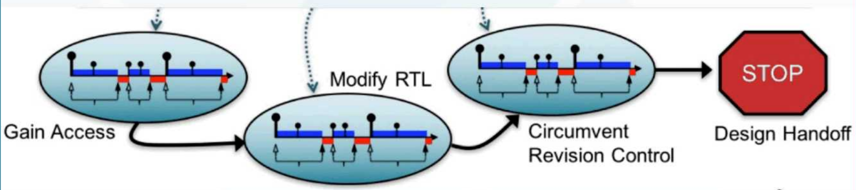
Threat Modeling Efforts

G-PLADD: Graph-based, Probabilistic Learning Attacker and Dynamic Defender*

Specified Attack Graph, Strategies



Series of Abstracted Games:
Attacker & Defender Interactions



Outputs: attack success probability, time to success, attack/defense costs, defender mitigations effectiveness, etc.

Strengths:

- Rapid evaluation of lots of attacks
- Representation of temporal attacker-defender interactions
- Adaptive, intelligent agents

Challenges:

- Input parameter development
- Abstract formulation limits ability to represent some attack specifics
- Requires additional effort for validation

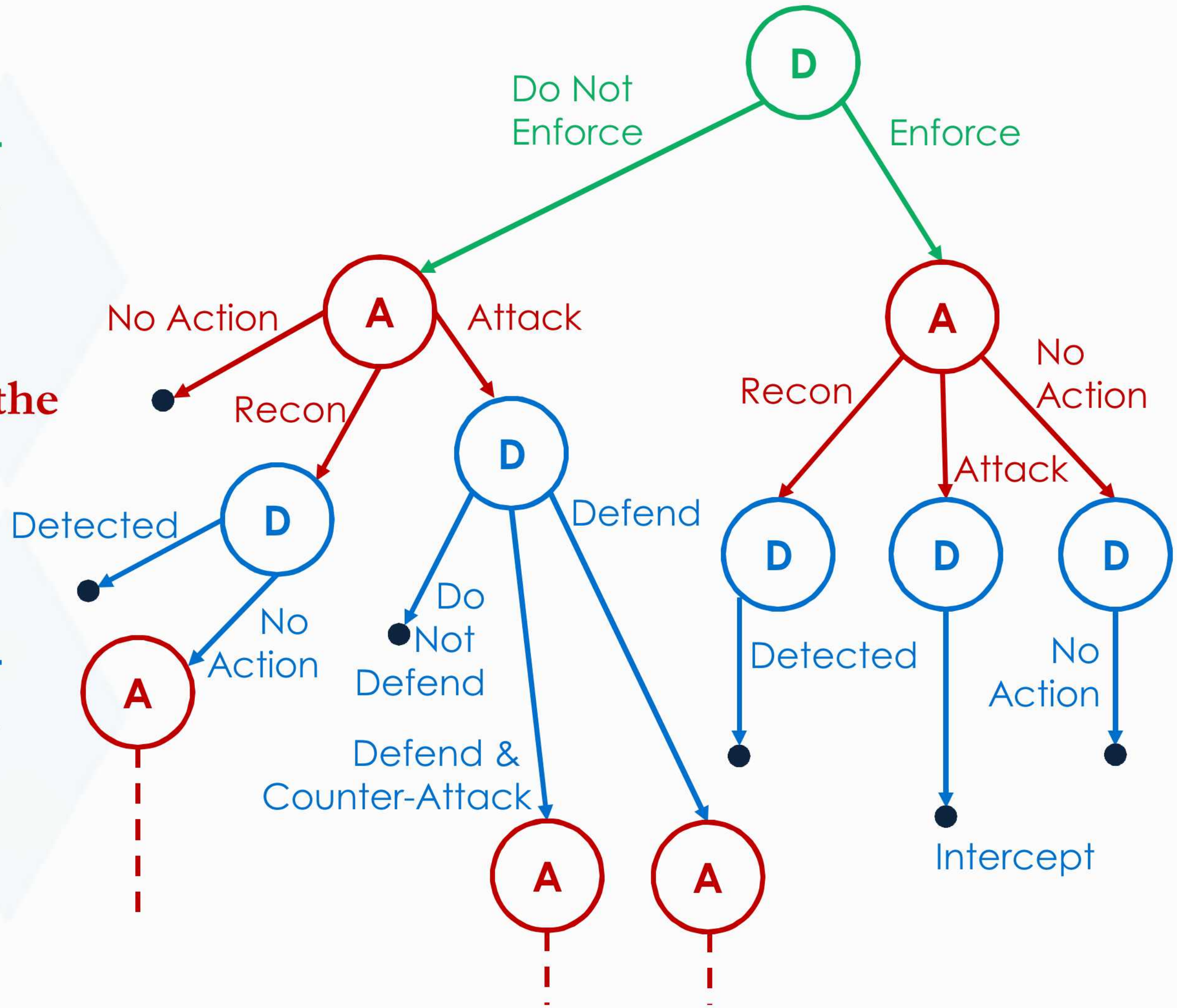
Multi-Stage Interdiction

An entity operates a cyber-enabled infrastructure and takes certain measures to defend it.

A cyber adversary attacks the entity to cause service disruption and physical damage.

An entity operates a cyber-enabled infrastructure and takes certain measures to defend it.

⋮



A New Class of Optimization Problems

Linear Programs

- Easily solved
- Widely used commercial solvers

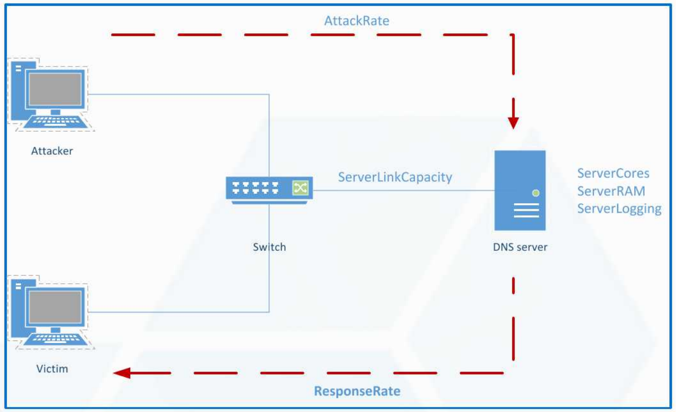
$$
\begin{array}{ll}
\min_{x \geq 0} & c^T x \\
\text{s.t.} & Ax \leq b
\end{array}
$$

Linear Bilevel Programs

- Hard problems (NP-hard)
- No general-purpose commercial solvers

$$
\begin{array}{ll}
\min_{x \geq 0} & c_1^T x + d_1^T y \\
\text{s.t.} & A_1 x + B_1 y \leq b_1 \\
 & \begin{array}{ll} \min_{y \geq 0} & c_2^T x + d_2^T y \\ & A_2 x + B_2 y \leq b_2 \end{array}
\end{array}
$$

Research thrust: Model confidence



?

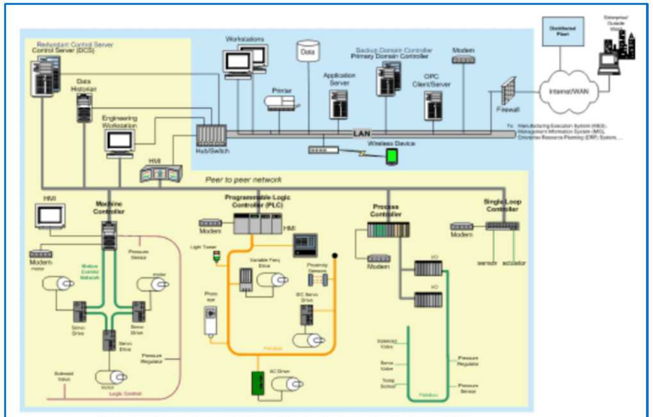


Figure 1: Generic Industrial Control System Network Architecture - DCS

From: Hieb, J., J. H Graham, and B. Luyster, *A Prototype Security Hardened Field Device for Industrial Control Systems.* 2019.

Question: How do we confidently make a V&V case at scale?

- **V&V Thrust 1**: Understand which uncertainties most affect model V&V
  - Collaboration with Kate Davis at Texas A&M
    - RESLab experiments on larger scale ICS systems
- **V&V Thrust 2**: Represent added complexity using coarser-grained models and assess convergence

UQ Support of Validation for Emulytics Models

- **Validation:**
  - Fundamental question: "Is this Emulytics model acceptable for this application?"
    - What level of network aggregation is acceptable?
    - Which quantities of interest should be used to make meaningful comparisons?
    - What are the validation metrics?

  - **Compare QoI distributions from Emulytics with Physical System**
  - **Compare QoI sensitivities from Emulytics with Physical System**

  - For small systems, Emulytics tools can be validated through *direct comparison* with experiments on actual networks.
  - As complexity increases, we will verify convergence in the sense that uncertainties and discrepancies *decrease* as more data and fidelity is added to the Emulytics model.

Results

Multi-fidelity modeling - setup

Network Configuration

- 1 client - 1 server (possible to extend to multiple clients)
- 100 Requests

Uncertain Parameters

- `DataRate` $\sim \mathcal{U}(5, 500)Mbps$
- `ResponseSize` $\sim \ln \mathcal{U}(500, 16 \times 10^6)B$

Fidelity definition

- `minimega` – HF: 100 Requests (average over 10 repetitions)
- `ns3` – LF: 10 Requests (`Delay` $50ms$)
- `ns3` – LF*: 1 Requests (`Delay` $5ms$)

	$\mathcal{C}$
HF	1
LF	0.016
LF*	0.002

TABLE: Normalized Cost

We assume **serial execution for the low-fidelity model**, however we might easily increase the efficiency of LF (`ns3`) by running multiple concurrent evaluations

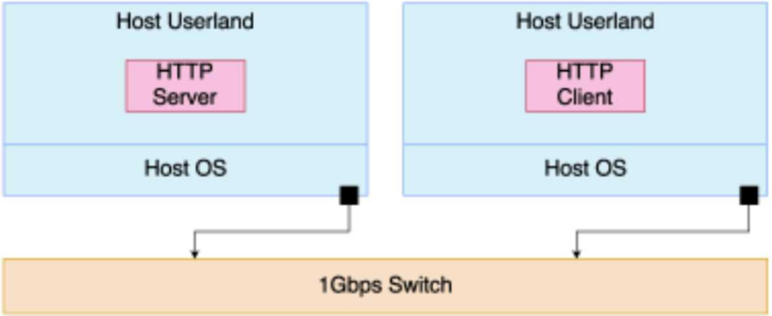


FIGURE: Network Configuration

Multi-fidelity modeling results – variance reduction

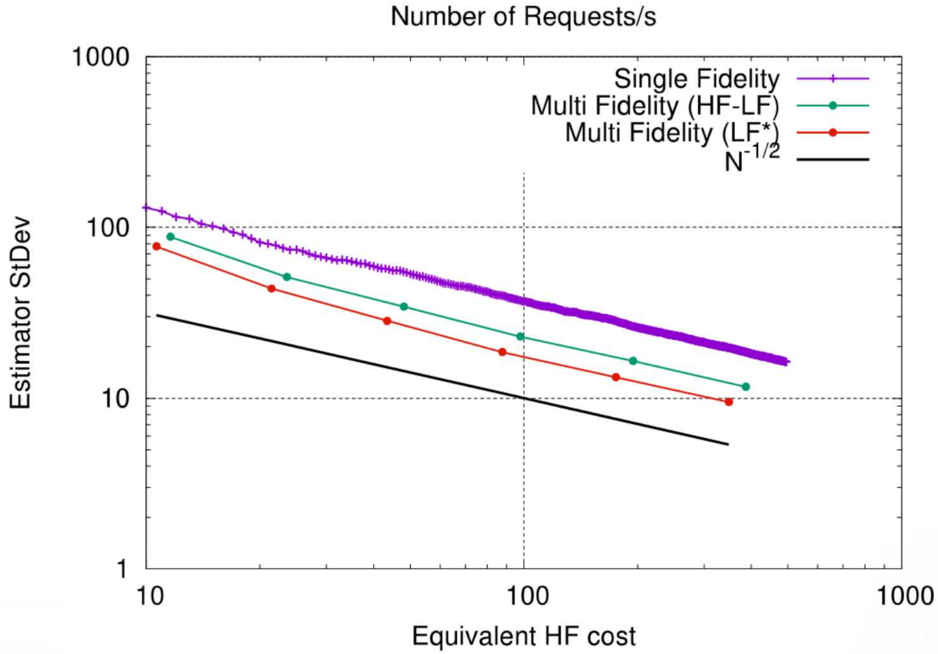


FIGURE: Exp. Value StDev

Example (for LF*)

- Number of **HF runs**: $N = 500$
- Number of **LF*** **runs**: $r_1 \times N = 5415$
- Equivalent **LF cost**: $r_1 \times N \times \frac{\mathcal{C}_{LF}}{\mathcal{C}_{HF}} = 11$
- **Total** estimator **cost** (HF + LF*): $\mathcal{C}_{tot} = 500 + 11 = 511$
- **Variance reduction**: $\left(1 - \frac{r_1 - 1}{r_1}\rho_1^2\right) = 0.23$

- The **variance reduction** we obtain w.r.t. MC is

$$\mathbb{V}ar\left(\tilde{Q}\left(\underline{\alpha}^{ACV}\right)\right) = \mathbb{V}ar\left(\hat{Q}\right)\left(1 - \frac{\mathbf{r_1 - 1}}{\mathbf{r_1}}\rho_{\mathbf{1}}^{\mathbf{2}}\right)$$

- The **number of low-fidelity simulations** is $N_{LF} = N \times r_1$ where

$$r_1 = \sqrt{\frac{\mathcal{C}_{HF}}{\mathcal{C}_{LF}}\frac{\rho_1^2}{1 - \rho_1^2}}$$

- For each HF simulation we need to spend an **extra cost** in LF simulations

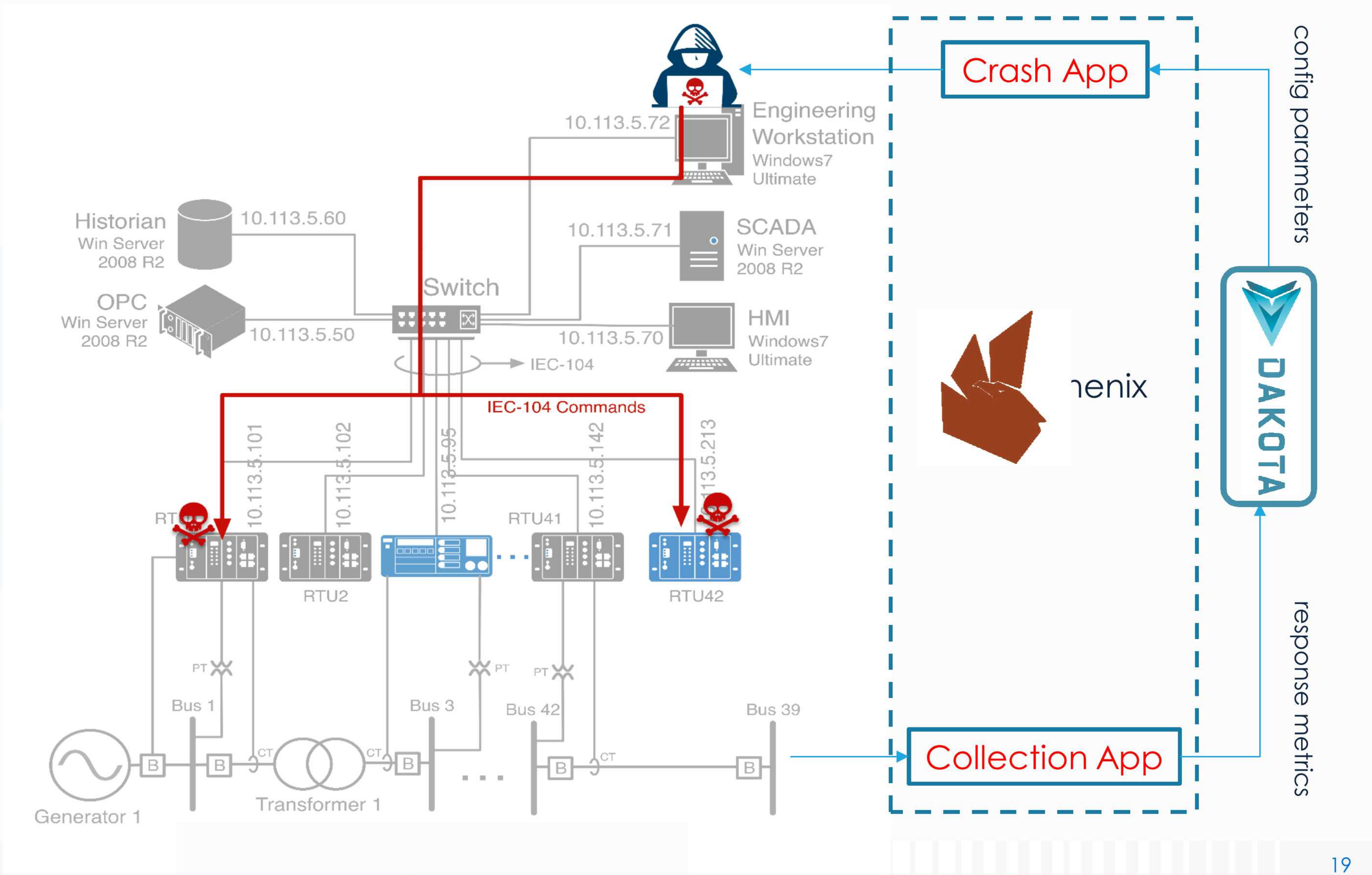
$$\text{Eq.Cost}: \quad \mathcal{C}_{tot} = N\left(1 + \mathbf{r_1}\frac{\mathcal{C}_{\mathbf{LF}}}{\mathcal{C}_{\mathbf{HF}}}\right)$$

- For this case

	ρ_1	r_1	$r_1\mathcal{C}_{LF}/\mathcal{C}_{HF}$
LF	0.86	4.69	0.075
LF*	0.90	10.83	0.022

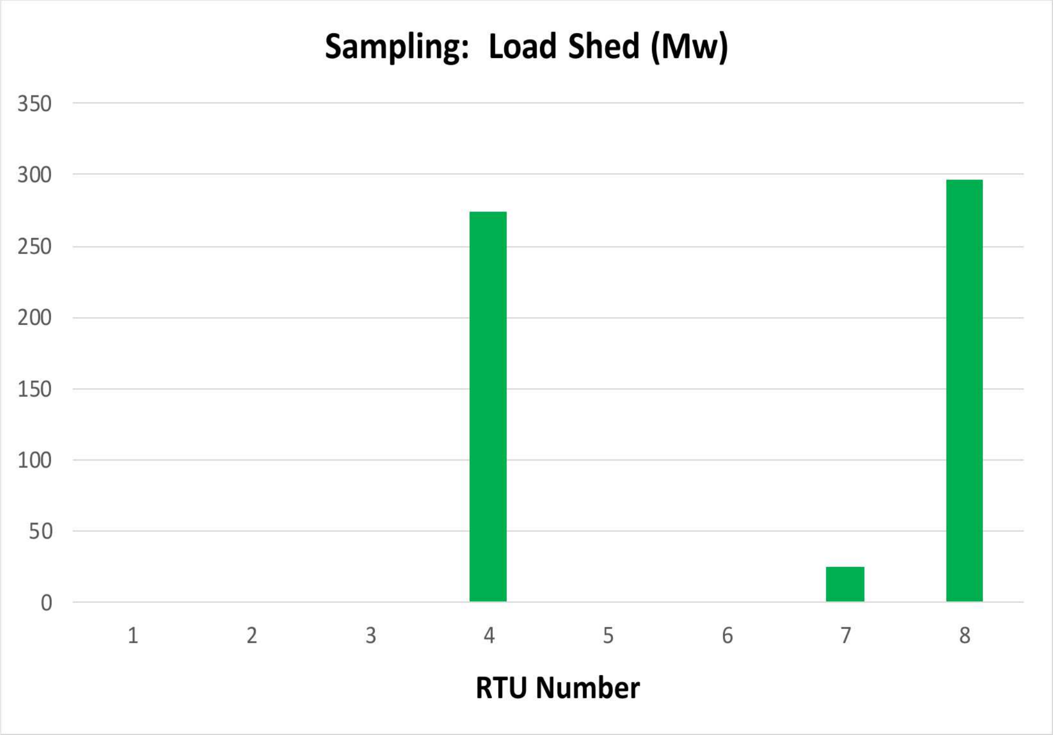
More than 70% variance reduction is obtained by adding **only an equivalent cost of 11 HF runs.**

SCEPTRE – DAKOTA Integration



Results: Impacts on Varying Target RTU

- There is variation in load shed when we target one RTU at a time
- Only three of the RTUs (#4, 7, and 8) generate effects on the response metric
- Results indicate that RTU-8 is a high-priority RTU for protection (followed closely by RTU-4)
- Given a limited budget, defender should not prioritize RTUs 1, 2, 3, 5, and 6



*Derived from synthetic data with no relation to actual grid: https://electricgrids.engr.tamu.edu/electric-grid-test-cases/activsg2000/

EXEMPLAR: Critical Component Identification

$$\max_{\delta\in\{0,1\}^{|\mathcal{R}|},w\in\{0,1\}^{|\mathcal{L}|},v\in\{0,1\}^{|\mathcal{K}|}} \gamma(\boldsymbol{\delta},\boldsymbol{w},\boldsymbol{v})$$

ATTACKER OBJECTIVE: Maximize Load Blackouts

subject to

$$\sum_r M_r \boldsymbol{\delta_r} \le M$$

Attacker's Budget

$$\sum_{r\in\mathcal{R}_k\cup\mathcal{I}_k}(1-\boldsymbol{\delta_r}) - |\mathcal{R}_k\cup\mathcal{I}_k| + 1 \le \boldsymbol{v_k} \le (1-\boldsymbol{\delta_r}), \quad \forall k\in\mathcal{K}, r\in\mathcal{R}_k\cup\mathcal{I}_k$$

$$\sum_{r\in\mathcal{R}_l\cup\mathcal{I}_l}(1-\boldsymbol{\delta_r}) - |\mathcal{R}_l\cup\mathcal{I}_l| + 1 \le \boldsymbol{w_l} \le (1-\boldsymbol{\delta_r}), \quad \forall l\in\mathcal{L}, r\in\mathcal{R}_l\cup\mathcal{I}_l$$

RTU Mapping to Physical Devices

$$\gamma(\boldsymbol{\delta},\boldsymbol{w},\boldsymbol{v}) = \min \sum_{g\in\mathcal{G}}\sum_{b\in\mathcal{B}} \boldsymbol{p_b^{L,S}}$$

DEFENDER OBJECTIVE: Minimize Load Disruptions

subject to

$$\boldsymbol{p_k} = \boldsymbol{v_k} B_k(\boldsymbol{\theta_{o(k)}} - \boldsymbol{\theta_{d(k)}} - \Theta_k), \quad \forall k\in\mathcal{K}$$

$$\sum_{g\in\mathcal{G}_b}\boldsymbol{p_g^G} + \boldsymbol{p_b^{L,S}} - \sum_{k\in\{k'|o(k')=b\}}\boldsymbol{p_k} + \sum_{k\in\{k'|d(k')=b\}}\boldsymbol{p_k} = \sum_{l\in\mathcal{L}_b} P_l^L, \quad \forall b\in\mathcal{B}$$

$$-S_k^{\max} \le \boldsymbol{p_k} \le S_k^{\max}, \quad \forall k\in\mathcal{K}$$

$$P_g^{G,min} \le \boldsymbol{P_g^G} \le P_g^{G,max}, \quad \forall g\in\mathcal{G}$$

$$\sum_{l\in\mathcal{L}_b}\boldsymbol{w_l}P_l^L \le \boldsymbol{p_b^{L,S}} \le \sum_{l\in\mathcal{L}_b}P_l^L, \quad \forall b\in\mathcal{B}$$

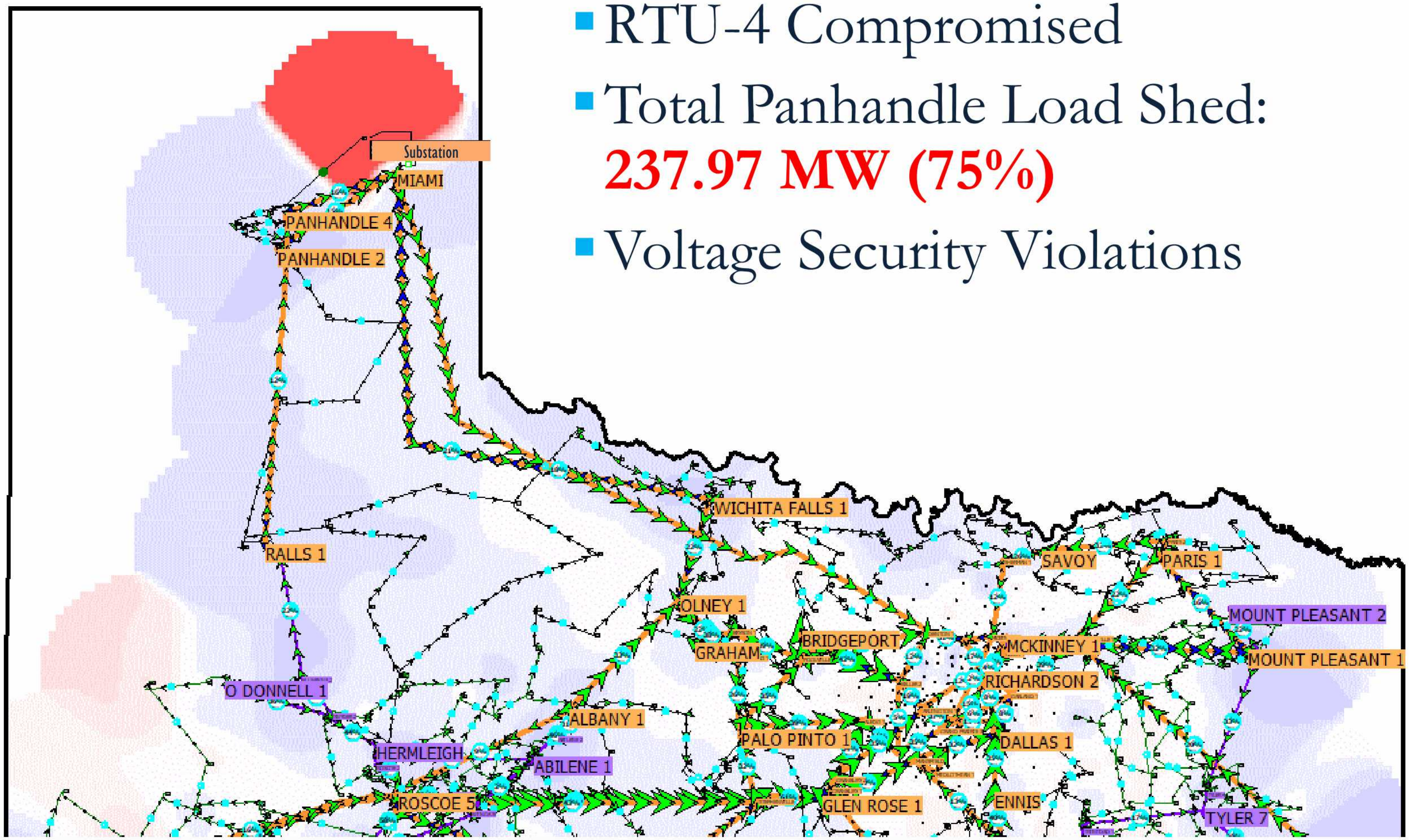
$$-\pi \le \boldsymbol{\theta_i} \le \pi, \quad \forall k\in\mathcal{K}$$

Steady-State Grid Operations

EXEMPLAR: Worst-Case RTU Attack

Attack Budget of '1':

- RTU-4 Compromised
- Total Panhandle Load Shed: **237.97 MW (75%)**
- Voltage Security Violations

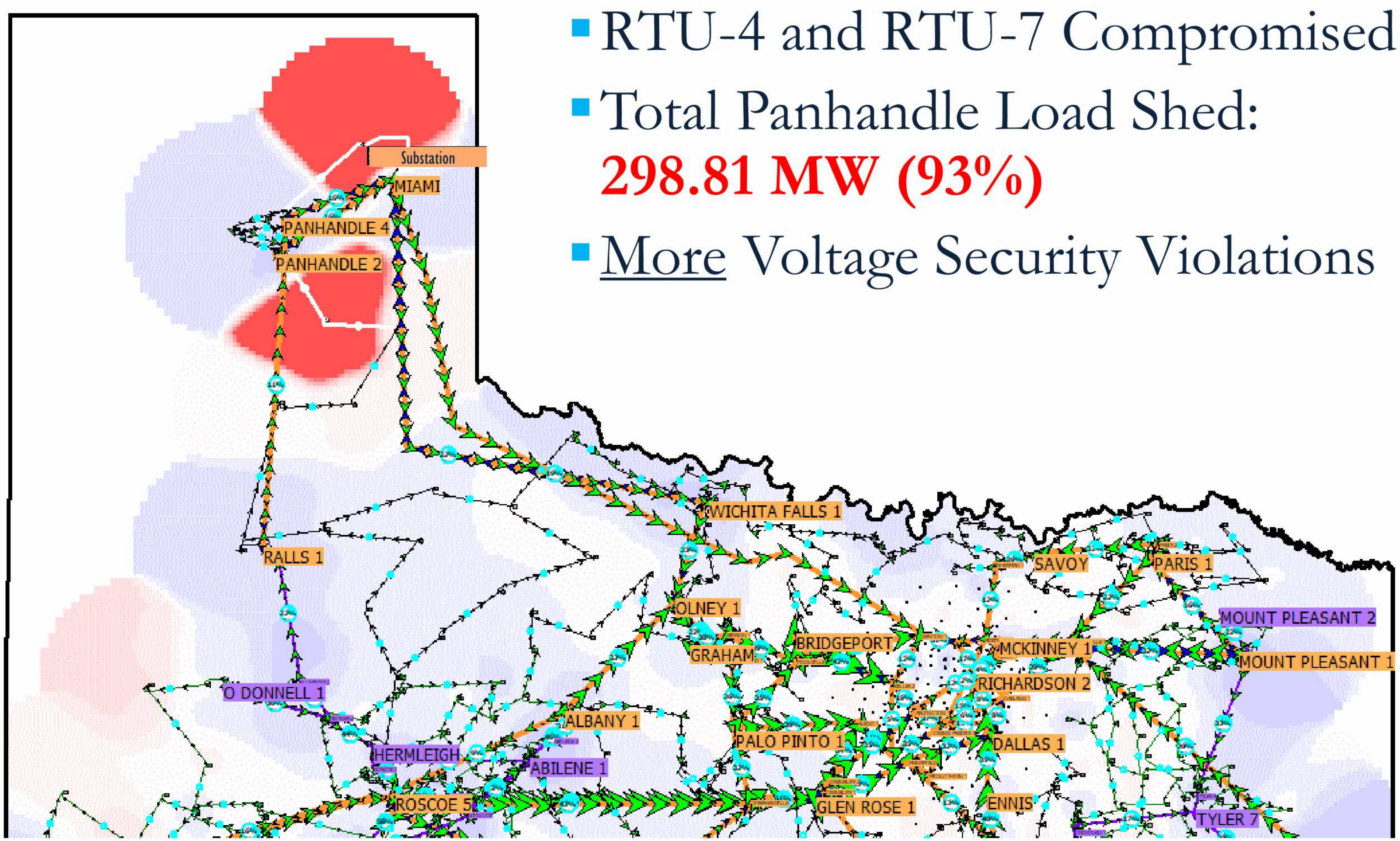


*Derived from synthetic data with no relation to actual grid: https://electricgrids.engr.tamu.edu/electric-grid-test-cases/activsg2000/

EXEMPLAR: Worst-Case RTU Attack

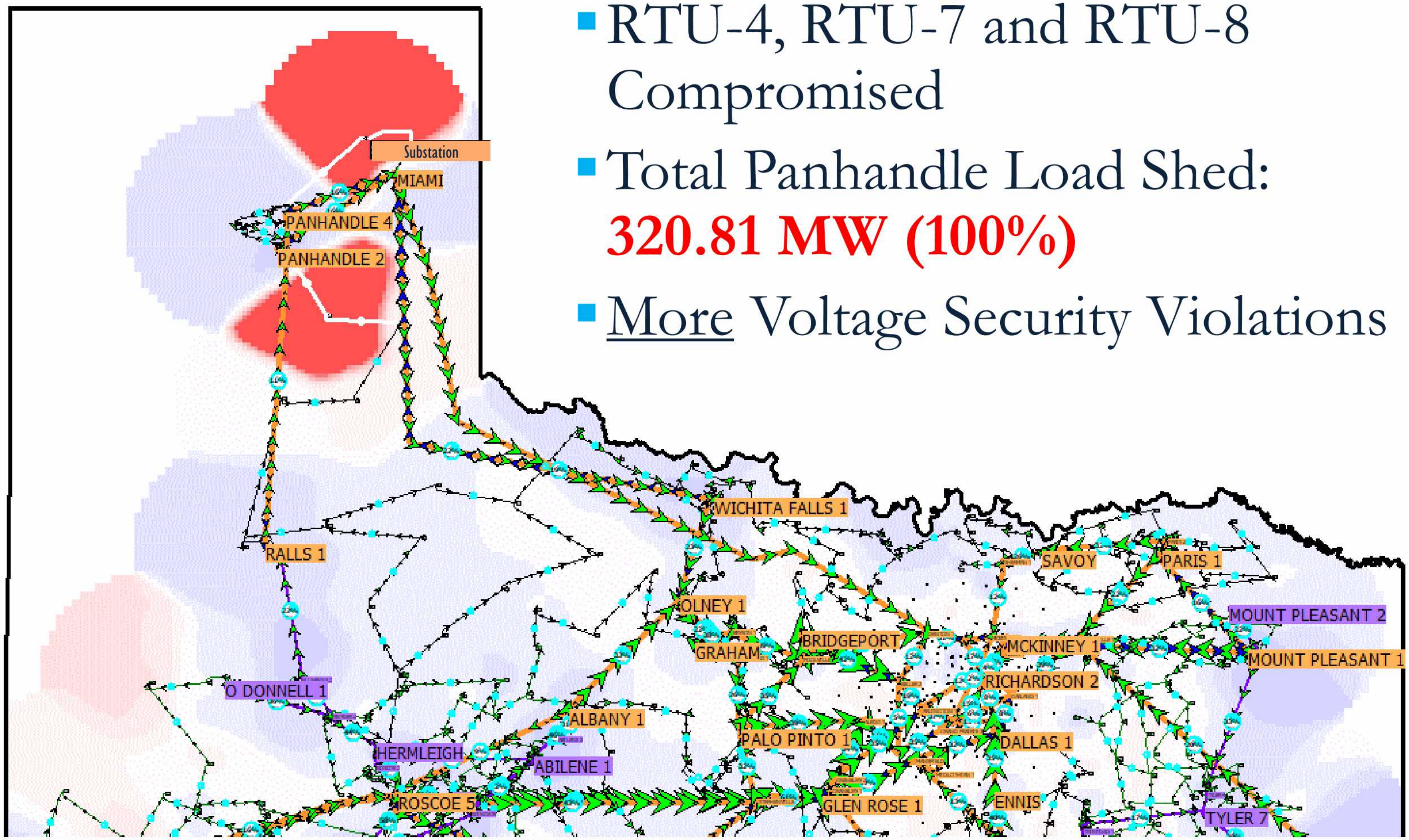
Attack Budget of '2':

- RTU-4 and RTU-7 Compromised
- Total Panhandle Load Shed: **298.81 MW (93%)**
- More Voltage Security Violations



*Derived from synthetic data with no relation to actual grid: https://electricgrids.engr.tamu.edu/electric-grid-test-cases/activsg2000/

EXEMPLAR: Worst-Case RTU Attack



Attack Budget of '3':

- RTU-4, RTU-7 and RTU-8 Compromised
- Total Panhandle Load Shed: **320.81 MW (100%)**
- More Voltage Security Violations

*Derived from synthetic data with no relation to actual grid: https://electricgrids.engr.tamu.edu/electric-grid-test-cases/activsg2000/

Summary

- Carefully design cyber experiments to:
  - Produce comprehensive, rigorous results
    - Needed for decisions about high consequence systems
    - Uncertainty quantification
  - Efficiently compute experimental iterations
    - State space explosion makes comprehensive coverage impossible
    - Dimension reduction, careful sampling to reduce the space
    - Optimization, game theory to identify regions of interest
    - Multi-fidelity modeling to generate statistics and reduce variance
- Capture uncertainty in threat
  - Use threat frameworks to track the threat
  - Use game theory and optimization formulations to determine:
    - Attack distributions for UQ
    - Worst case threats
    - Best defense strategies
- Rigorously construct a validation case
  - Use uncertainty quantification to identify sensitive parameters and responses
  - Assess convergence when adding
    - Fidelity (e.g. physical experiments)
    - Data (e.g. additional runs, real-world data, etc.)