SAND2021-6332PE

LDRD

Laboratory Directed Research and Development

Rigorous Cyber Experimentation for Science of Security

Ali Pinar

Sandia National Laboratories

apinar@sandia.gov

U.S. DEPARTMENT OF ENERGY

NNSA National Nuclear Security Administration

Sandia National Laboratories is 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-NA0003525.

UNCLASSIFIED UNLIMITED RELEASE

SECURE: **S**cience and **E**ngineering of **C**ybersecurity by **U**ncertainty quantification and **R**igorous **E**xperimentation

- SECURE aims to develop cyber experimentation techniques to
  - answer “what if questions” with high-confidence **(Emulytics)**
  - assess confidence in results under uncertainty (**Uncertainty Quantification)**
  - make robust decisions under uncertainty in an adversarial environment (**Adversarial Optimization)**

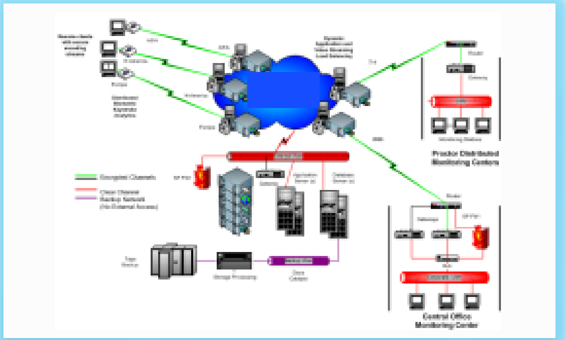
with rigor.

- Lack of rigor limits adoption for high-consequence decisions

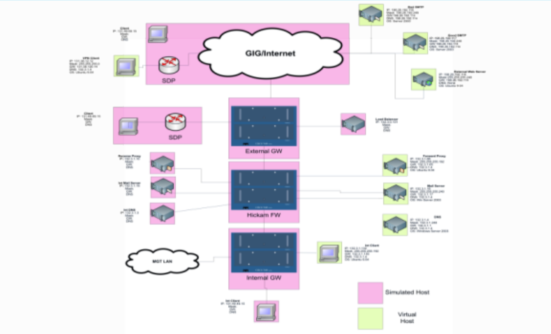
- The cyber experimentation process is analogous to the scientific method
  Hypothesis → test → analyze results → repeat

- SECURE’s success will advance cyber experimentation to be a pillar of science of cybersecurity
  - similar to computational science and engineering for physics based systems

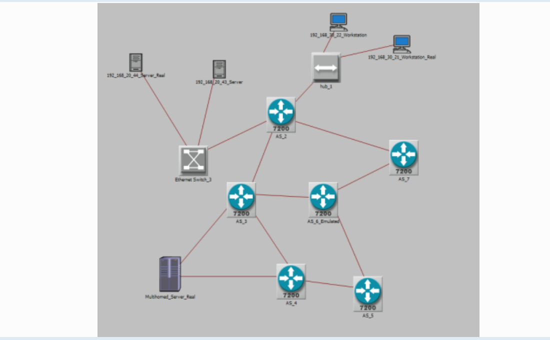
Cyber experimentation approaches



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

SECURE's approach:

- Results should be independent of the platform and the tools used for the experiment
- Multi-fidelity techniques enable utilizing advantages of multiple methods

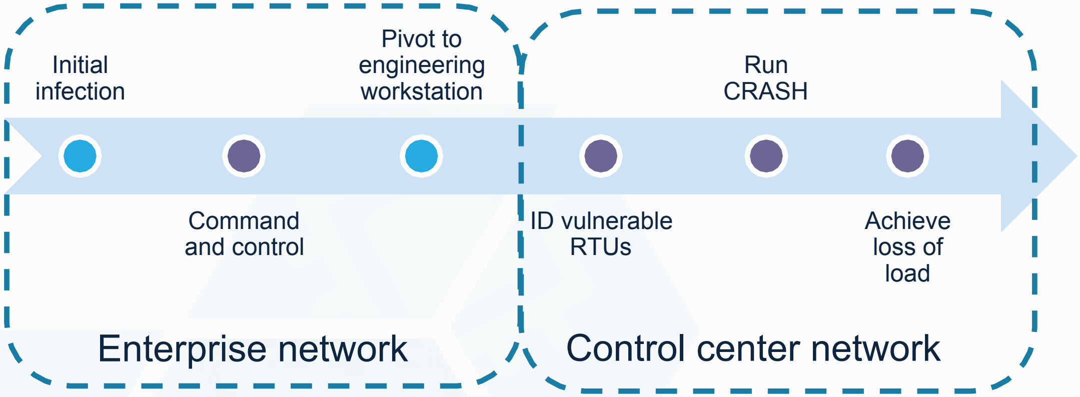
An Overview of the Process

Question: Is our power grid resilient against an attack as in Ukraine?

- Ukraine attack was based on Crash Override Malware
- The attacker gains remote access to power grid components to turn them on and off.

- Cyber Experimentation Process:
  1. Model the attack
  2. Model the cyber system and its defenses
  3. Model the consequences of the attack
  4. Find remediations

Attack the Model

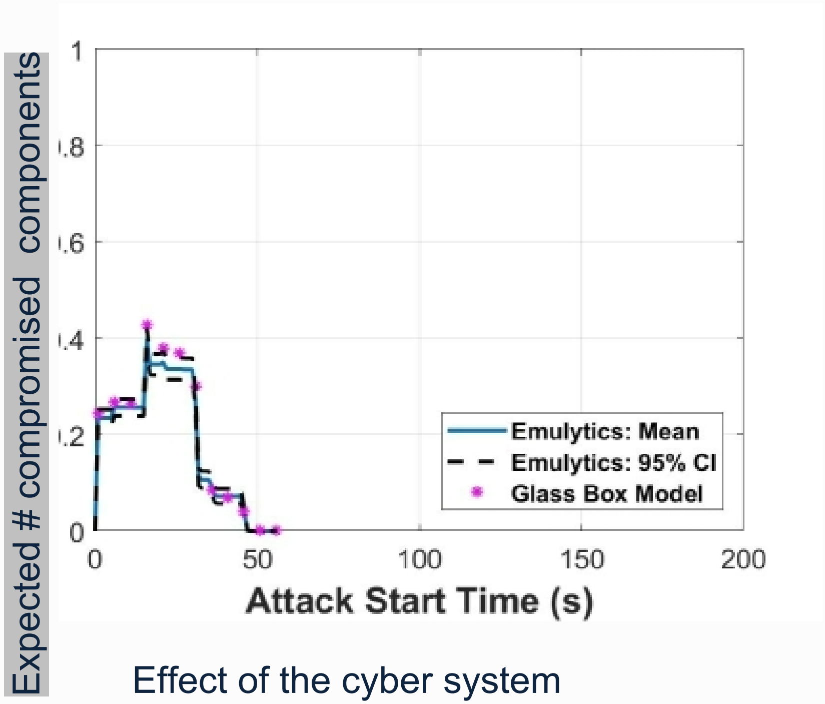


- Need to model the attacker capabilities in a parameterized way
- Attack databases have data
  - We transform data into information

Research Challenges:
How do we represent knowledge?
How do we customize a model for specific system?
How do we quantify an attacker’s success probability?

Model the Cyber System

- Build a model of the cyber system and apply the attack
- Run many scenarios
- Analyze the data

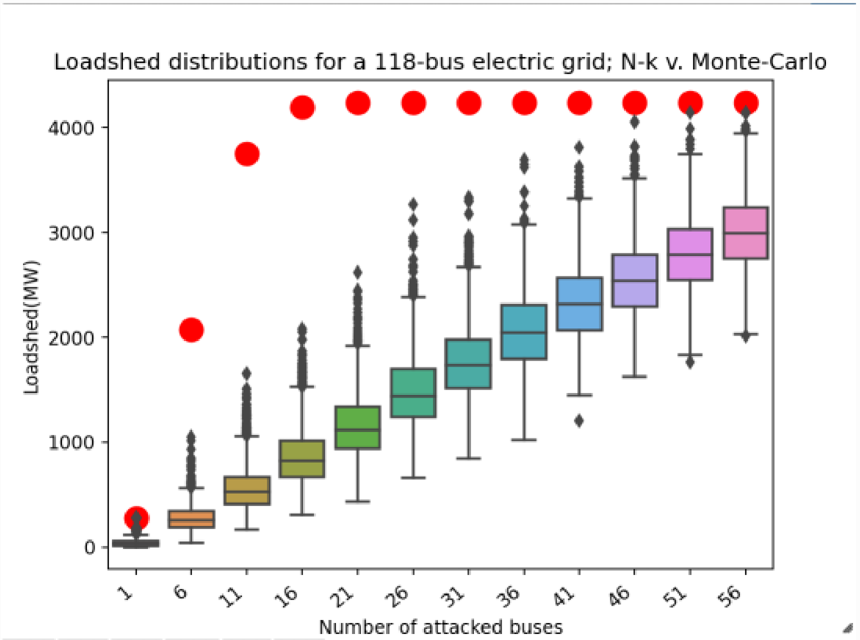


Effect of the cyber system

Research Challenges:
Verification and Validation
Model input uncertainties
Scenario orchestration
Models I with varying fidelities
Uncertainty propagation in high dimensions
Scalability

Analyze the Consequences

- The main goal is resilience of the system being supported
- Having understood the effect on the cyber system, we investigate effect on the physical system

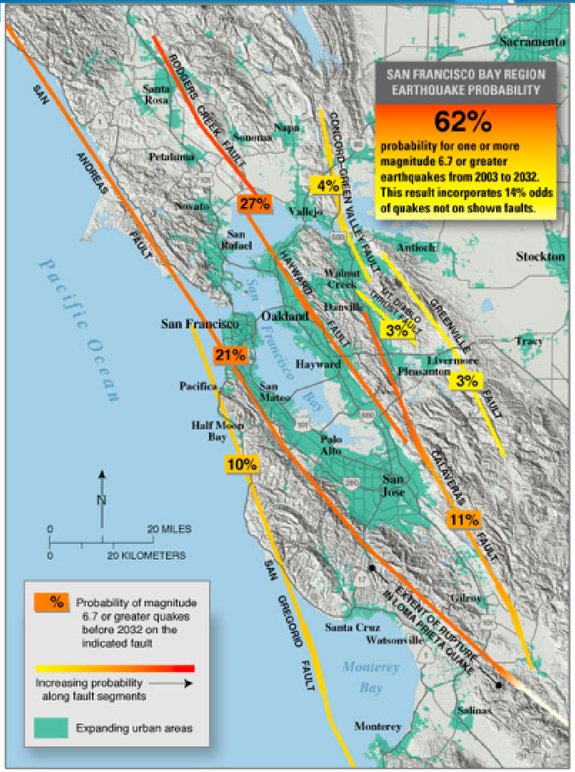


Loss of load on the power system

Research Challenges:
Extreme events
Tail probabilities
Scalability
General purpose adversarial optimization solvers
Model validation

Remediation: Cyber-aware resilience and Consequence-aware cyber defense

- How do we improve cyber-systems for better resilience?
  - ○ Attacks equivalent in cyber metrics lead to different consequences
  - ○ Current work: network segmentation
- How do we operate on physical systems in a cyber threat-informed way?
  - ○ What is a cyber fault line?
  - ○ Current work: cyber-aware attack models



California Fault Lines

Research Challenges:

What is a good cyber/physical interface?

How do we design systems that are resilient by design?

How do we deal with increasing uncertainty for full system assessment?

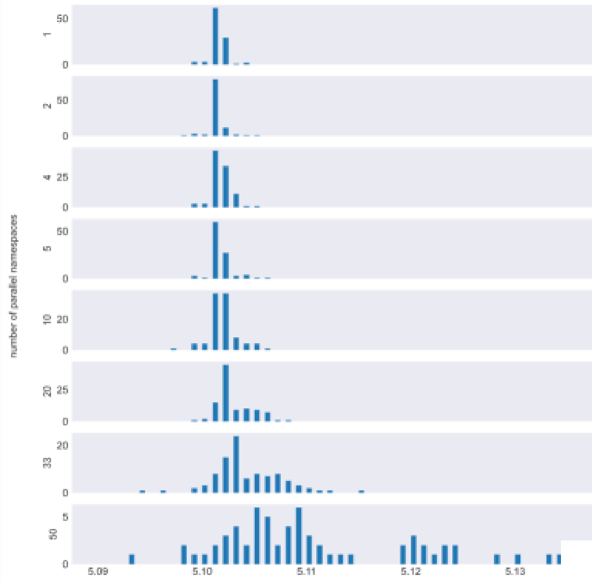
How do we identify sensitive parameters in discrete/high-dimensional systems?

Genesis of SECURE is to investigate Verification and Validation of cyber experiments at scale

- Build on:
  - V&V concepts from the computational science community
- A few core ideas:
  - Verification: Are equations solved correctly?
    - Software quality: unit testing, regression testing, etc.
    - Numerical analysis, stability, convergence analysis.
  - Validation: Is the model adequate to use for the intended application?
    - Quantitative comparison between experiment (physical test) and model.
    - Accounts for uncertainties and errors in both experimental data and model.
- Adaptation for Emulation:
  - Verification: Do virtual machines operate in environment with proper realism?
  - Validation: How do we measure adequateness at scale given randomness in experiments?

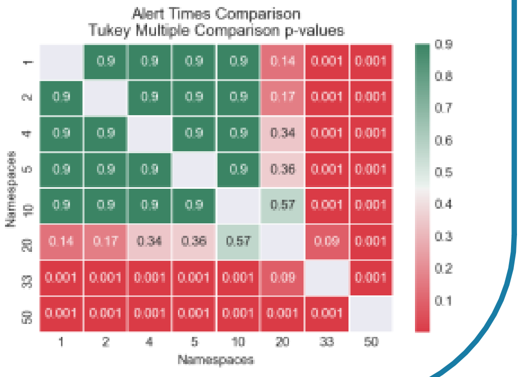
Verification: Effect of adding too many namespaces

- Distribution of alert times shift as namespaces are added
- Quantified similarity with Tukey Multiple Comparison Test
  - Shows clear drop in similarity after 10 namespaces
- Large p-value indicates that the null hypothesis can't be rejected
  - $H_0: \mu_1 = \mu_2$
  - **Larger p-value -> similar results**



Alert Times Distribution

Tukey Multiple Comparison


Alert Times Comparison
Tukey Multiple Comparison p-values

Namespaces	1	2	4	5	10	20	33	50
1		0.9	0.9	0.9	0.9	0.14	0.001	0.001
2	0.9		0.9	0.9	0.9	0.17	0.001	0.001
4	0.9	0.9		0.9	0.9	0.34	0.001	0.001
5	0.9	0.9	0.9		0.9	0.36	0.001	0.001
10	0.9	0.9	0.9	0.9		0.57	0.001	0.001
20	0.14	0.17	0.34	0.36	0.57		0.09	0.001
33	0.001	0.001	0.001	0.001	0.001	0.09		0.001
50	0.001	0.001	0.001	0.001	0.001	0.001	0.001	

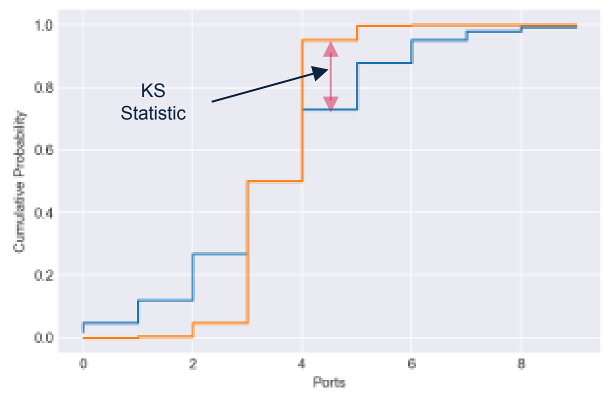
Research Challenges:

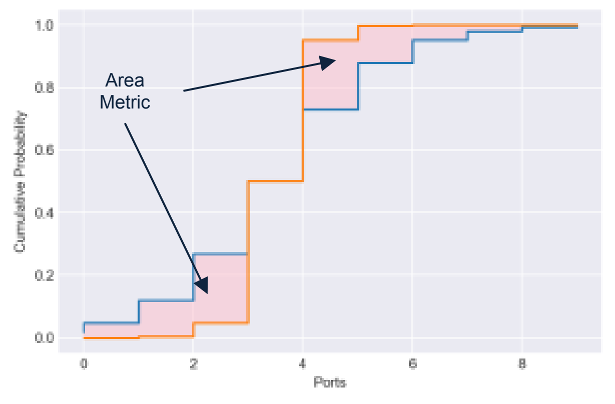
What are the hardware invariants that can indicate system overload?

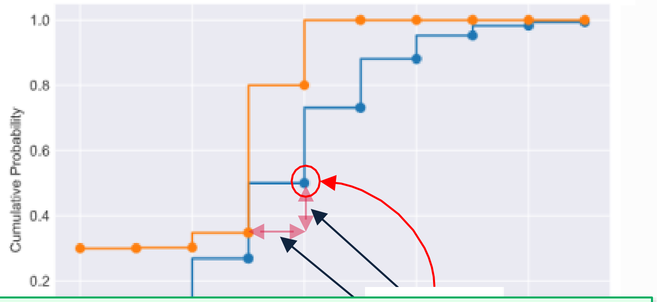
How do we measure efficiently? How do we analyze (in-situ)?

Validation: Comparing results

- **Area Metric**
  - Sum of the differences in area between the CDFs of two samples [1]
  - This is not a p-value, **small values imply similarity**





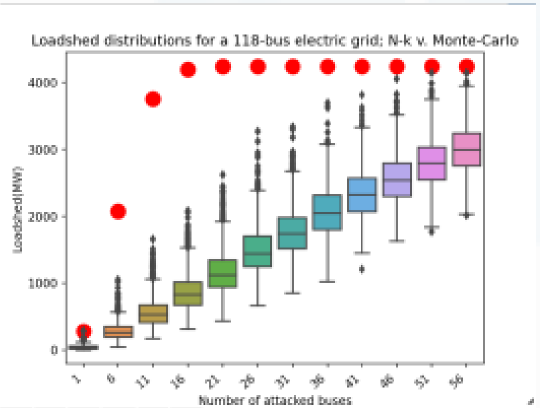


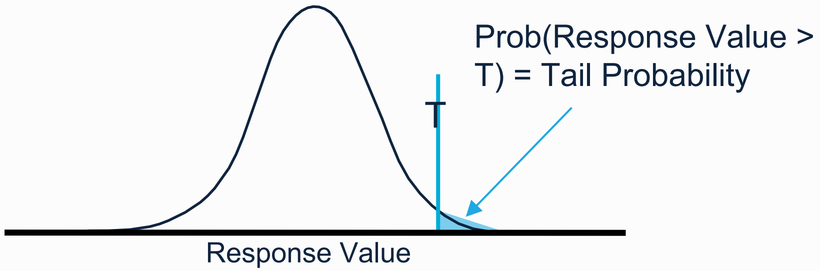
[1] K.A. Maupin, L.P. Swiler, N.W. Porter, “Validation Metrics for Deterministic and Probabilistic Data,”

Research Challenges:
How do we model noise? What are proper metrics/ time scales for comparison?
How do we scale algorithms? How do we build representative smaller systems?

Always/Never Systems

- We need to identify events with low-likelihood yet high-consequence
  - Solution: Multi-fidelity sampling for tail events; optimization for extreme points





Prob(Response Value > T) = Tail Probability

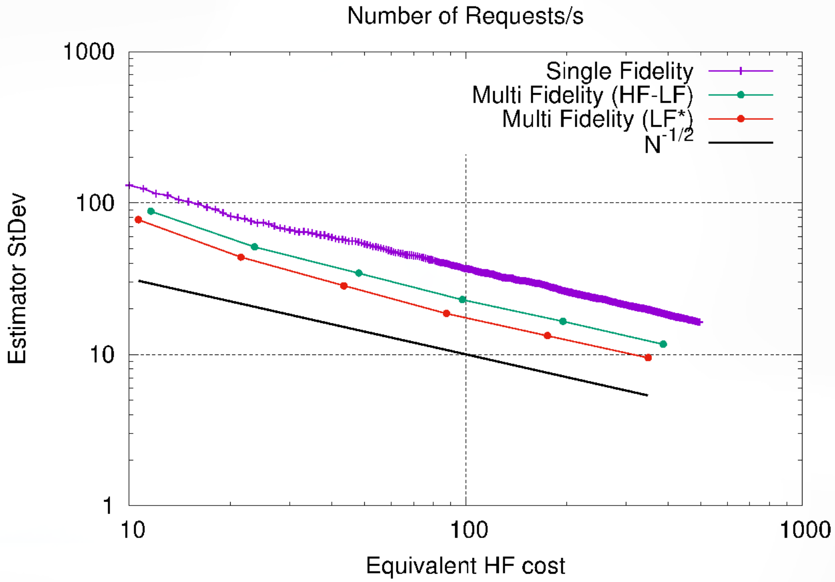
T

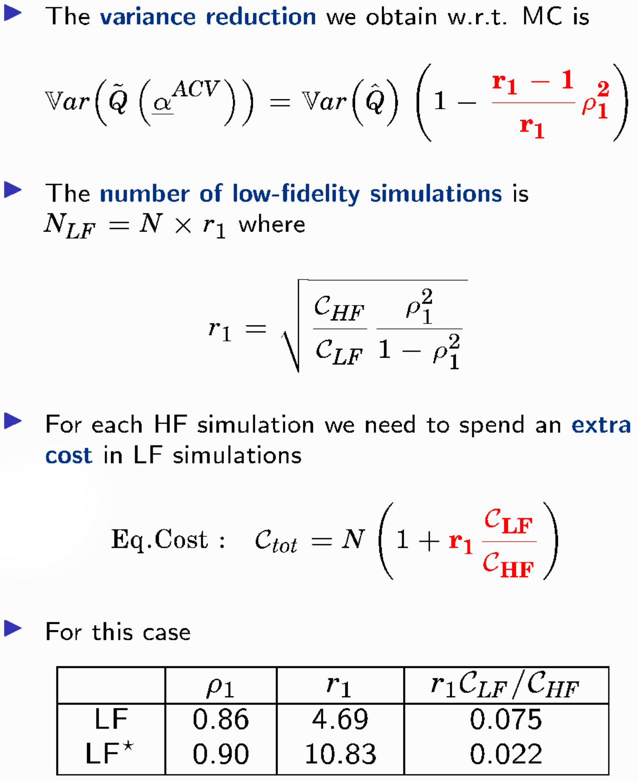
Response Value

- We need to face the sparse heterogenous data problems
  - High-fidelity data will be limited; we need to work with multi levels of fidelity.
  - Solution: Multi-fidelity methods use a small number of high-fidelity model runs (emulation) augmented with many lower fidelity runs (simulation or mathematical models) to reduce the variance in the results. This requires correlation between the high and low fidelity models.

Multi-fidelity modeling results – variance reduction

- 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 two models





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

$$\mathbb{V}ar\left(\tilde{Q}\left(\underline{\alpha}^{ACV}\right)\right) = \mathbb{V}ar\left(\hat{Q}\right)\left(1 - \frac{\mathbf{r_1 - 1}}{\mathbf{r_1}}\rho_1^2\right)$$

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

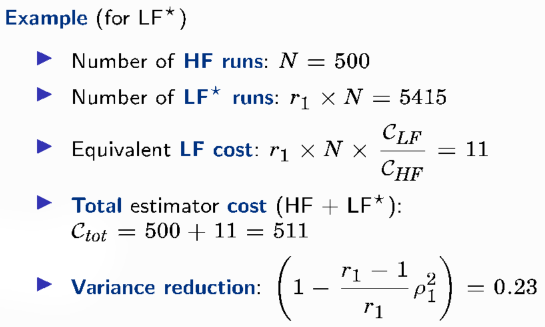
$$r_1 = \sqrt{\frac{\mathcal{C}_{HF}}{\mathcal{C}_{LF}}\frac{\rho_1^2}{1-\rho_1^2}}$$

- 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



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**: $\left(1 - \frac{r_1 - 1}{r_1}\rho_1^2\right) = 0.23$

. Value StDev

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

Adversarial Optimization

Linear Programs

- Easily solved
- Widely used commercial and academic solvers

$$
\begin{array}{ll}
\min_{??\geq 0} & ??^{??}\, ?? \\
\text{s.t.} & ????\leq ??
\end{array}
$$

NOTE: These methods are not cyber or grid specific

Linear Bilevel Programs

- Hard problems (NP-hard)
- No general-purpose commercial solvers for **discrete lower level decisions**

$$
\begin{array}{ll}
\min_{??\geq 0} & ??_1^{??}\, ?? + ??_1^{??}\, ?? \\
\text{s.t.} & ??_1\, ?? + ??_1\, ?? \leq ??_1 \\
 & \begin{array}{ll} \min_{??\geq 0} & ??_2^{??}\, ?? + ??_2^{??}\, ?? \\ & ??_2\, ?? + ??_2\, ?? \leq ??_2 \end{array}
\end{array}
$$

Upper Level Problem

Lower Level Problem

Conclusions

- Cyber experimentation can be a pillar of science of cybersecurity
- Technology for cyber experimentation is advanced,
  - But needs to be supported mathematical tools to apply scientific principles
- SECURE is leading the way,
  - Made significant progress but still long way ahead
    - In depth and in breadth
  - Many opportunities for collaboration

- Our success will
  - Provide decision support for high-consequence systems
  - Design systems of the future that can be resilient to anticipated threats
  - Compare solutions in realistic settings
  - Quantify security, and thus the return on investment, in a principled way
  - Present a capability for
    - Prediction and data generation for extreme events
    - Inference for model generation when data is sparse

Acknowledgments

- The work is supported by the Laboratory Directed Research and Development program of Sandia National Laboratories.
- Many people contributed to this work.

Emulytics Team

- *Tom Tarman*
- *Jerry Cruz*
- *Eric Vugrin*
- *Meg Sahakian*
- *Seth Hanson*
- *Sasha Outkin*
- *Tim Schulz*
- *Christian Reedy*

UQ Team

- Laura Swiler
- Bert Debusschere
- Gianluca Geraci
- Trevor Rollins
- Erin Acquesta
- Jon Crussell

Optimization Team

- *Jared Gearhart*
- *Bill Hart*
- *Anya Castillo*
- *Bryan Arguello*
- *Cindy Phillips*
- *Emma Johnson*
- *She'ifa Punla-Green*