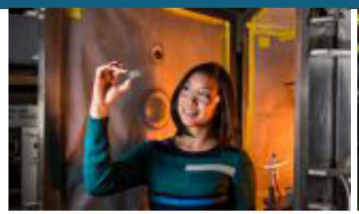
SAND2019-14156PE

Sandia National Laboratories

SECURE LDRD Grand Challenge External Advisory Board Review #2 Charge to the EAB



PRESENTED BY

Heidi Ammerlahn, Director
Champion

October 29 - 30, 2019

LDRD

ENERGY NNSA

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.

EAB Charter

The External Advisory Board (EAB) to the Science and Engineering of Cybersecurity by Uncertainty quantification and Rigorous Experimentation (SECURE) LDRD Grand Challenge (GC) will serve in an advisory capacity to assess and provide external, independent review and guidance to the GC team and management at Sandia National Laboratories on the project's Strategy, Relevance/Impact, Quality, Capabilities, and Partnerships. *(See next slide)*

The EAB will provide feedback on these five elements of Research Assessment throughout the three-year course of the project. Focus questions highlighting particular elements of interest will be provided at each meeting, but the EAB is encouraged to comment on any or all elements, and to provide additional advice to the project team as desired.

Five Elements of Research Assessment

ELEMENT	
Strategy	Execute a research strategy that is clear, aligns discretionary investments (e.g., LDRD) with the research strategy, and supports DOE/NNSA priorities.
Mission Relevance & Impact	Ensure that research is relevant, enables the national security missions, and benefits DOE/NNSA and the nation.
Quality	Ensure that research is transformative, innovative, leading edge, high quality, and advances the frontiers of science and engineering.
Capabilities	Maintain a healthy and vibrant research environment that enhances technical workforce competencies and research capabilities.
Partnerships & Technology Transitions	Research and develop high-impact technologies through effective partnerships and technology transfer mechanisms that support the laboratory's strategy, DOE/NNSA priorities, and impact the public good.

External reviews are linked to
SNL Performance Objective: Science, Technology and Engineering Mission

SECURE GC EAB #2 Focus Questions

(Strategy) At its first meeting in March 2019, the EAB recommended that SECURE should map out a single project architecture, clearly define what comprises success for individual tasks and the project as a whole, and stake out integration activities to be accomplished throughout the project.

- Please provide feedback on how effectively we communicate (a) the rationale ("story") of the project and (b) the research plan – is it sound, comprehensive, executable? What adjustments to the overall plan or individual thrust areas should we consider?
- What are the biggest technical / programmatic risks in Year 2 of the project, and what changes would you recommend to address them?

The exemplar for Year 2 is focused on placement of malware on a SCADA network and subsequent consequences to the power grid. Is this an appropriate exemplar; what enhancements would you suggest? Specifically:

- What degree of progress, if any, do you see with respect to optimization, UQ, scalability, and validation?
- How effectively does the exemplar demonstrate the propagation of these attacks in a probabilistic manner and how they can be optimally mitigated? How well does it illustrate an effective connection between the thrusts?

(Quality) Is the research demonstrated to date of high quality and at the leading edge of the cyber experimentation community? What, if anything, needs to be sharpened / improved?

(Partnerships & Technology Transition) Please comment on the team's plans for engagement with external communities and "Life after LDRD" (e.g. identification of sponsors who will adopt key accomplishments / support further development.)

AGENDA

Day 1: October 29, 2019

Director Champion Welcome, Comments, and Charge to the Board - 8:15-8:45

Project Rationale and Overview - 8:45-9:45, Break - 9:45-10:00

Emulytics - 10:00-11:30

Break and Pick up Food for Working Lunch 11:30-12:00

Uncertainty Quantification - 12:00-1:30, Break: 1:30-1:45

Optimization - 1:45-3:15

Thrust integration and longer term technical vision 3:15– 3:45

Break - 3:45-4:00, EAB Closed Session - 4:00-5:00, Quick questions / feedback 5:00 – 5:15

Dinner for EAB Members, PI, PM, Director Champion, and LDRD Office 6:30 – 8:00

Day 2: October 30, 2019

Address questions from the previous day - 8:30-8:45

Programmatic Vision / Life after LDRD discussion – 8:45-9:45

Charge Review/Concluding Questions - 9:45-10:00

Break 10:00-10:15

EAB Closed Session - 10:15-11:30

EAB Outbrief to Team - 11:30-12:30

Adjourn at 12:30; some EAB members remain for other discussions in the afternoon

LDRD

Laboratory Directed Research and Development

SECURE Overview

Ali Pinar, PI

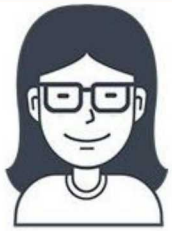
Zach Benz, PM

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.

What is the return on investment for cyber security?

Credit: Staff Sergeant Jason Gamble, United States Air Force

Who else cares about return on investment?



Adison the Engineer, IT decision maker
"Will deploying this cybersecurity solution have meaningful impact?"

Captain Howard, DoD, high-consequence systems
"Can we credibly assess system performance under various threat scenarios?"

Alice the Architect, Grid resilience planning
"How do we take into account cyber vulnerabilities in grid operations?"
"How do we take into account consequences in cyber system design? "

Leon the PM and Dr. Turing the PI, capability stewards
"What are the gaps in our capability roadmap to focus on to maximize impact? "

Olivia G. Arcane, Government systems analyst
"Which part(s) of our system is most fragile?"

evidence

Without ~~data~~,
you are just another person with an opinion.

W. Edwards Deming

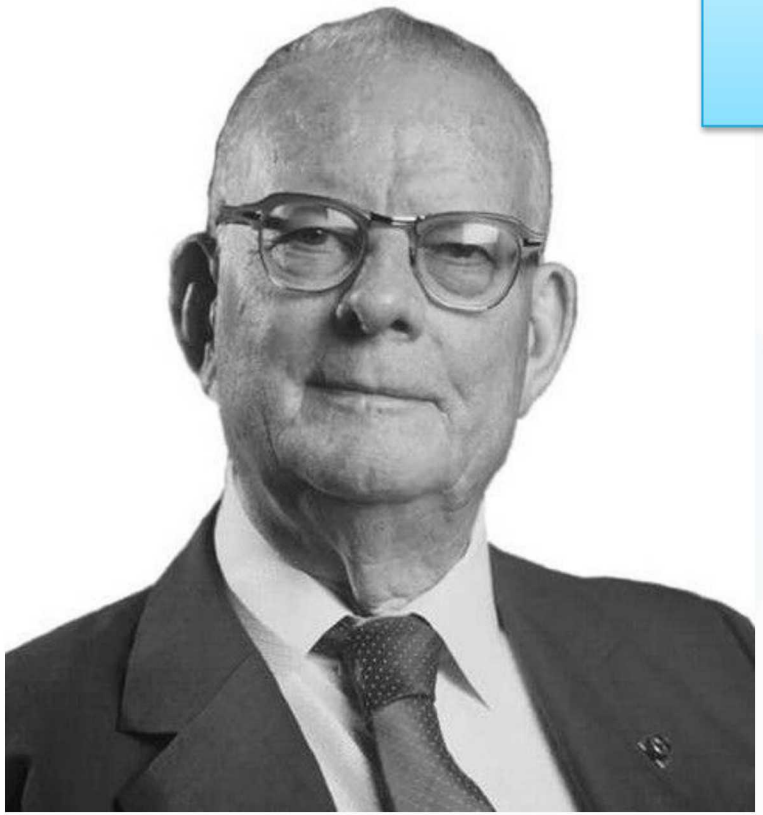
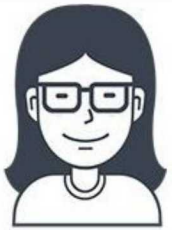


Image Source: census.gov

Why do we need to quantify?



Adison the Engineer, IT decision maker

"Will deploying this cybersecurity solution have meaningful impact?"

- Her team offers different opinions about the potential benefits of the proposed solution and its impact on productivity
- She needs a thorough cost/benefit analysis to base her decision on

Captain Howard, DoD, high-consequence systems

"Can we credible assess system performance under various threat scenarios?"

- He is in charge of a high-consequence system
- He trusts his red team, but the stakes are too high; the system is too complex; and time is too short

Leon the PM, capability steward

"What are the gaps in our capability roadmap to focus on to maximize impact? "

- He controls a budget that is too small; needs to prioritize
- Many conflicting expert opinions; system is too complex for the answers to be simple

Bringing Rigor into Cyber Experimentation: The Plan in a Nutshell

SECURE: Science and Engineering of Cyber security through Uncertainty quantification and Rigorous Experimentation

The Goal: Bring rigor into cyber experimentation

The Idea: Follow the principles of Computational Science and Engineering (CSE)

The Challenge: Cyber systems are different than those in traditional CSE applications.

The Plan:

- Build on our current strengths in core capabilities
  - Emulytics, Uncertainty Quantification (UQ), Optimization
- Advance the state of the art in core capabilities
- Integrate core capabilities over a power grid exemplar

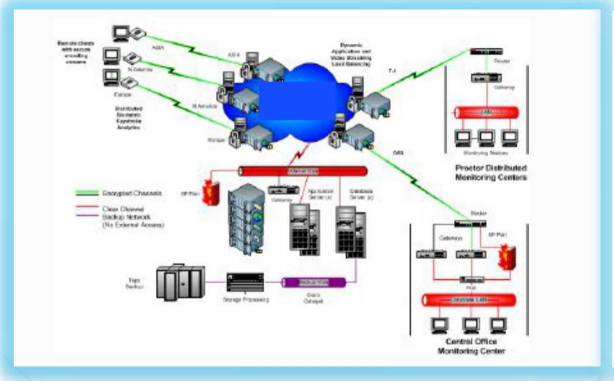
EMULYTICS

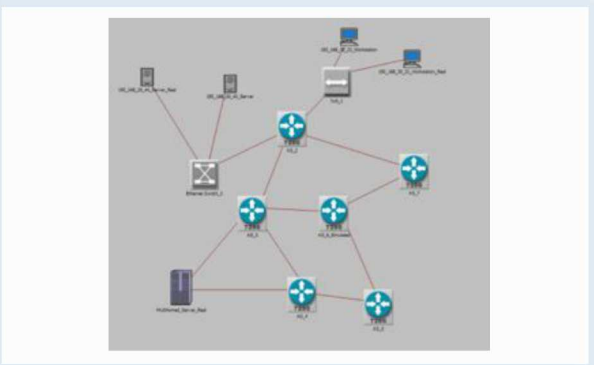
DAKOTA

PYOMO

The Product: Algorithmic expertise to support the full workflow of rigorous cyber experimentation and software tool SECUREtk

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

Challenge is bringing together disparate strengths

What we need

- Predict Answer "What if questions" at scale, with confidence.
  - Emulytics
- Assess confidence in predictions; characterize and propagate uncertainties
  - Uncertainty Quantification
- Make robust decisions under uncertainty and under advanced threat conditions
  - Adversarial Stochastic Optimization

Three Thrusts of SECURE

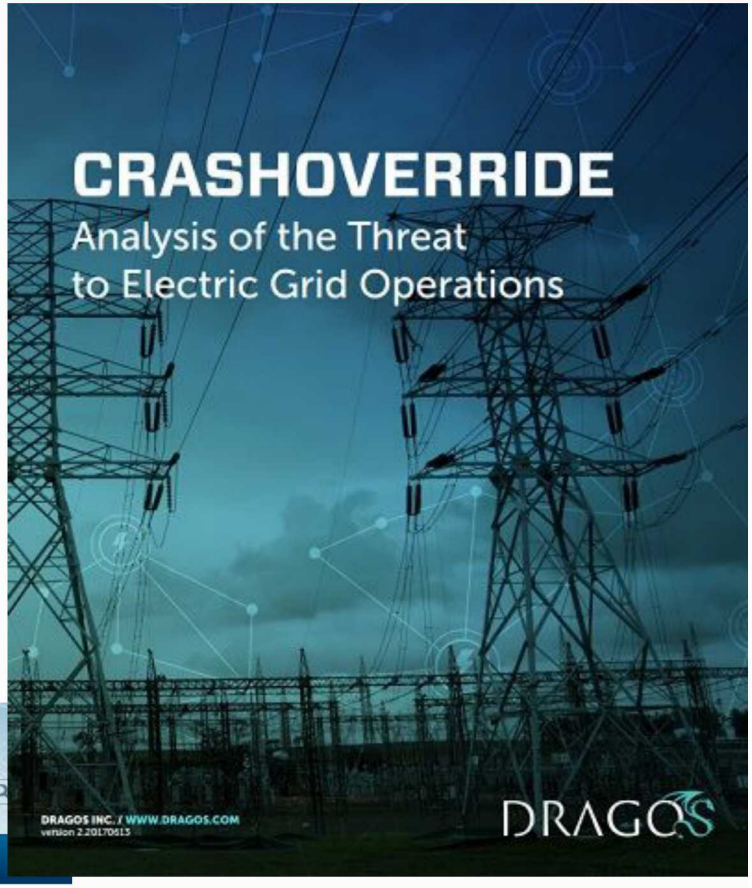
EMULYTICS

DAKOTA

PYOMO

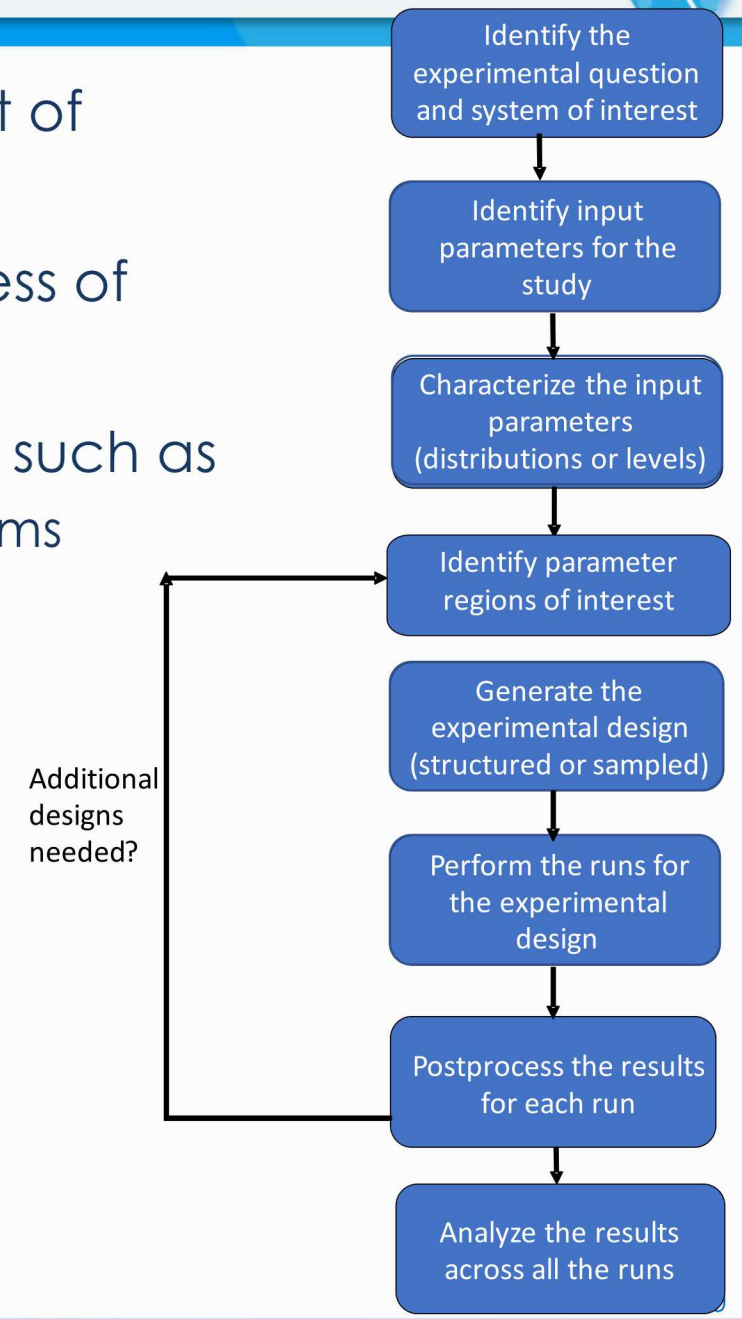
Three Research Elements and One Motivating Application to Tie Them All

- The Electric power grid is a cyber-physical system that is becoming increasingly information dependent.
- The 2015 Ukranian power grid attack showed the potential effects of a cyber attack on a critical infrastructure.
- This motivating application helps us better understand how pieces fit together.



What is the SECURE Product?

- A cohesive **experimental workflow** and set of **tools and techniques…**
- …that **rigorously quantifies** the effectiveness of actions in the cyber domain
- We will develop foundational capabilities such as
  - UQ for discontinuous, high dimensional systems
  - Scalable solvers for optimization
  - Scalable Emulytics
- .., and produce
  - Experts
  - Publications
  - Algorithms
  - Tools to share (SECUREtk, DAKOTA, Pyomo, minimega)



Overview of the Exemplar Study/Workflow

Threat Model

- Crashoverride on a single ICS
- Focus on part of the attack Reconnaissance
- Attacker needs to act quickly
- Attacker tries to locate RTUs using *nmap*
- Defender tries to detect such searches using *snort*
- Parameter ranges set for a fast strike attack

Attack Effect on Resources

- A representation of the a region of the Texas Grid
- Flat cyber network
- Controls 8 RTUs
- Build an emulation model of the system
- Run the emulation many times to cover the parameter space
- Build models for impact on cyber
- Validate models using emulation

Consequence Prediction

- Quantify impact on the power grid based on loss of load
- Investigate how a sophisticate adversary can use this attack in an optimal way
- Provide feedback to previous steps about sensitive parameters regions

Research Plan (Overview)

Year 1: Integration and Algorithmic Exploration

- Surveys; apply present capabilities; integration (tools and ideas); initial results for new ideas; fine-tuned problem definitions
- Exemplar 1: Single operating authority; flat SCADA/RTU network;
- Products: Prototype implementations; papers on early results; integrated experimental environment

Year 2: Algorithm Development

Deep dive into algorithmic research; testing at scale/complexity; research software; initial demonstration of new, joint capabilities
Exemplar 2: Regional; SCADA/RTU network; multiple ICS networks
Products: SECUREtk 0.1 (internal use); Algorithm publications

Year 3: Demonstrate Capability

Pushing the boundaries of tools; Reporting results; demonstration of capabilities; research software to tools;
Exemplar 3: Western Grid; IT/SCADA/RTU network; multiple subnets & services
Products: SECUREtk 1.0 (sharable with research partners); Integration publications

Overview of Year 1 Progress

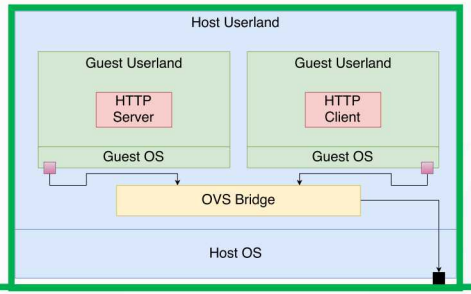
- Initial results that tie security to scientific foundations
- Detailed plan for the power grid exemplar
- Demonstration of integration of the exemplar
- Invited talks and publications
- IAB Review
- EAB Review
- Initial External Engagements with a wide customer space and research partners
- On schedule with all milestones
- Communications:
  - Started SECURE Seminar Series: 4 talks so far
  - Started Quarterly Newsletter: 1st issue out, 2nd issue in progress
  - Domains name: securegc.sandia.gov; sandia.gov/securegc

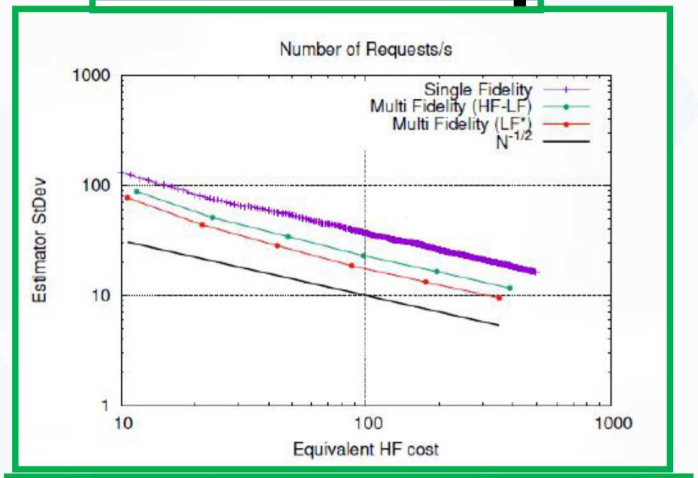
Integration has been the primary goal

- Goal: SECURE should be an **interdisciplinary team** that will create **foundational** technologies for rigorous **cyber-experimentation.**
- Our plan: Start with the integration and let the research grow out of the common roots
  - Conference room reserved for a full day for SECURE activities
  - Initiated flow of information between research elements in the first year
  - Developed common language
  - Avoided integration only through team leads; encourage individuals to understand other fields; build a network

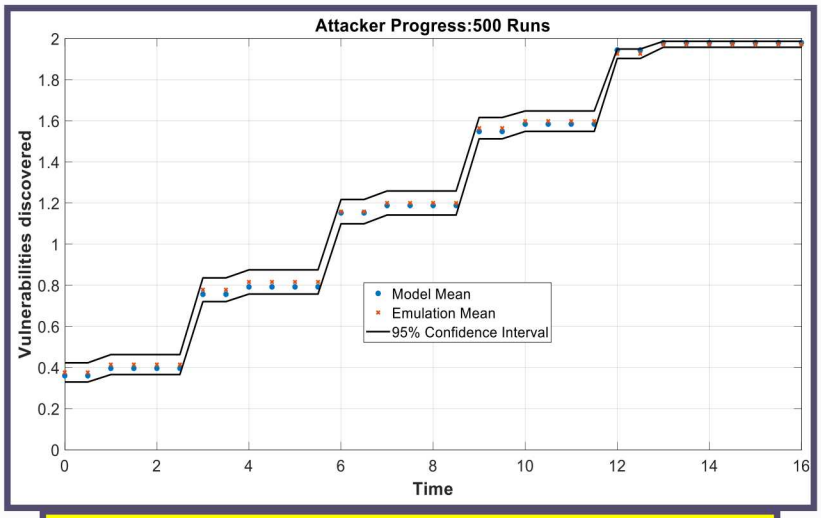
At the end of the year, we are a team that can
- develop interdisciplinary solutions
- ask questions we could not have asked before
- better understand the limits of current methods and fundamental challenges behind practical problems

Many Promising Early Results



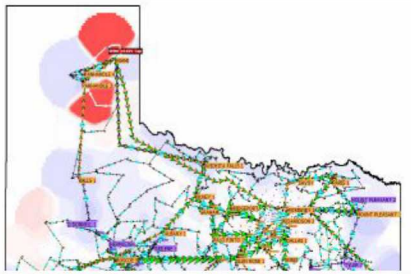


Adoption of Multi-fidelity UQ techniques for high- confidence models for communication times



A glass box model build from emulation
for threat characterization

Ability to identify worst case scenarios without enumeration
General purpose SW in progress



Multiple RTUs Compromised

*Derived from synthetic data that does not represent actual grid: https://electricgrids.engr.tamu.edu/electric-grid-test-cases/activsg2000/

Research Plan (Overview)

Year 1: Integration and Algorithmic Exploration

- Surveys; apply present capabilities; integration (tools and ideas); initial results for new ideas; fine-tuned problem definitions
- Exemplar 1: Single operating authority; flat SCADA/RTU network;
- Products: Prototype implementations; papers on early results; integrated experimental environment

Year 2: Algorithm Development

- Deep dive into algorithmic research; testing at scale/complexity; research software; initial demonstration of new, joint capabilities
- Exemplar 2: Regional; SCADA/RTU network; multiple ICS networks
- Products: SECUREtk 0.1(internal use); Algorithm publications

Year 3: Demonstrate Capability

- Pushing the boundaries of tools; Reporting results; demonstration of capabilities; research software to tools;
- Exemplar 3: Western Grid; IT/SCADA/RTU network; multiple subnets & services
- Products: SECUREtk 1.0 (sharable with research partners); Integration publications

FY20 Plan: Integrated Overview

- **Scaling up the Exemplar**
  - More complex SCADA networks; Multiple ICS networks;
  - More complicated questions; restructuring the network; tailored defenses
  - Regional attacks: attacks on multiple ICS
  - Detailed models for higher dimensions
- **Leveraging Exemplars to Showcase Use Cases of SECUREtk**
  - Risk management through stochastic adversarial optimization
  - Design of Emulytics experiments with analytical methods
  - Constructing confidence intervals under high response variability
  - Hypothesis testing to guide experimental design
- **Quantifying V&V as part of the Cyber Experimental Process**
  - Conducting V&V in the context of problem
  - Requirements analysis to assess well-posedness of cyber models
  - Extending methods to address boundary conditions and estimating tail probabilities
- **Integration of Threat Characterization to Experiment Ensembles**
  - Pruning meaningless and low-consequence attack spaces prior to running Emulytics experiments
  - Identifying optimal mitigation strategies that deter or evolve the threat space

EAB Feedback from March 2019

- Overall positive impressions
  - SECURE proposes to address a long-overdue research challenge …..
  - … proposed a sound approach …
  - … assembled a talented team that demonstrated ability to leverage experience gained form prev. projects
  - The overall plan for the research and the progress to date were impressive …
  - …quickly built strong cross-disciplinary collaborative relationships…
- with constructive feedback on
  - what was missing in the presentation
    - Publication list; CVs – new web page
    - Sponsorship targets – Zach's presentation on Day 2
    - Project architecture – end of the day presentation
    - Threat characterization – Tom's presentation today
  - what was missing in the project plan
    - ….

EAB Feedback

- Community awareness
  - Worked with Perspectives to explore the space broadly
  - Collaborations with Academia
    - GTech, Texas A&M, UC Davis, RPI, UC Berkeley
  - Talks/sessions in conferences
  - Workshop in the works
  - SECURE Speakers
- Power systems domain expertise
  - NE ISO visit; GTech and Texas A&M collaborations; on-team and local experts
- People development
  - New hires, new team members, new roles for team members
  - Graduate students
- Risks identification
  - Next slide

Risk Mitigation

- (Un)Realistic and (not) representative exemplar
  - Working with domain experts; avoid real data for classification; *focus on research*
- Cascading effects of a delayed task
  - If there is a delay, proceed with synthetic data
  - Defined our interfaces to ensure that an output can be rigorously computable.
  - So even if there is a delay, we are confident that integration is feasible
- Miscommunication issues in integration
  - This was the first task; we will integrate early and often
- Difficulty of validation
  - Start with small problems
  - Focus on methodology, so that we are ready when we have the data
- Uncertainty margins too large to be practical
  - Identify the source and improve if we can
  - If not, proving wide margins is useful.

March EAB feedback items

CA
- Community awareness
  - "Future presentations should explicitly acknowledge related research and articulate how SECURE is going beyond it. "

RI
- Risks identification
  - "explicitly identify risks associated with the project and develop strategies to mitigate them "

TC
- Threat characterization
  - "it was not clear to the board how the threat characterization work contributed to the overall goals of the project "

ST
- SECUREtk

DE
- Domain expertise
  - "given that the exemplar is the power grid, the EAB did not see sufficient evidence that the team has the required domain expertise to create realistic scenarios that will answer meaningful questions "

PA
- Project architecture
  - "The board suggests that the team map out a project "architecture" that shows how tasks are connected … [and]the team needs to clearly define what comprises success and stake out integration activities to be accomplished throughout each year of the project"

CE
- Customer engagement
  - "it was not apparent either who the specific customers will be for SECURE's output or that the research plan is appropriately addressing medium- to long-term customer challenges "

PD
- People development
  - "The EAB was unclear on SNL's development / promotion of talent and expertise in cybersecurity"

LDRD

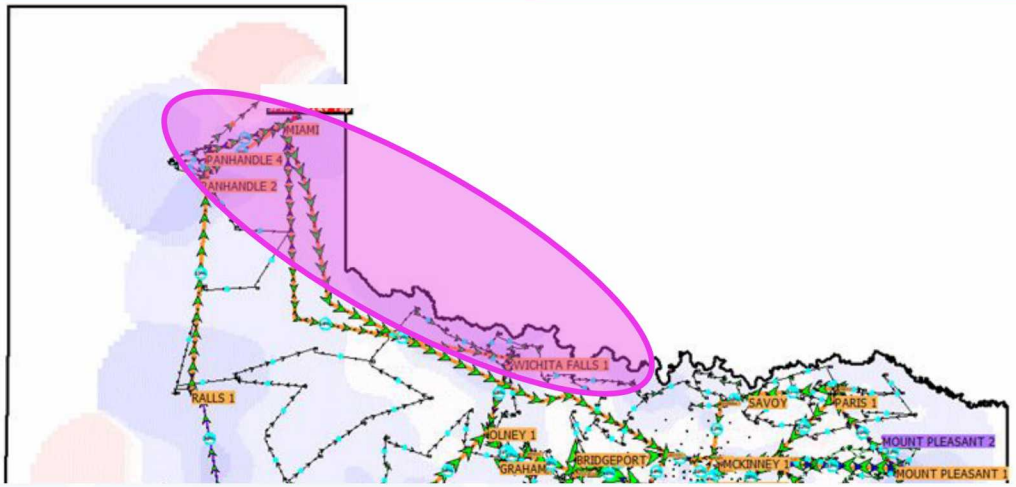
Laboratory Directed Research and Development

Simulated Case Study: Scenario Description

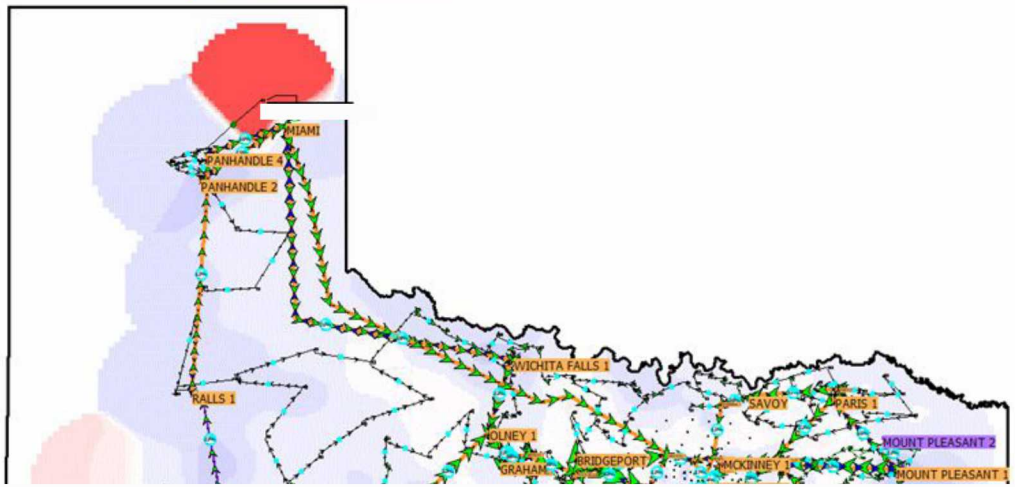
Presenter: Eric Vugrin

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.

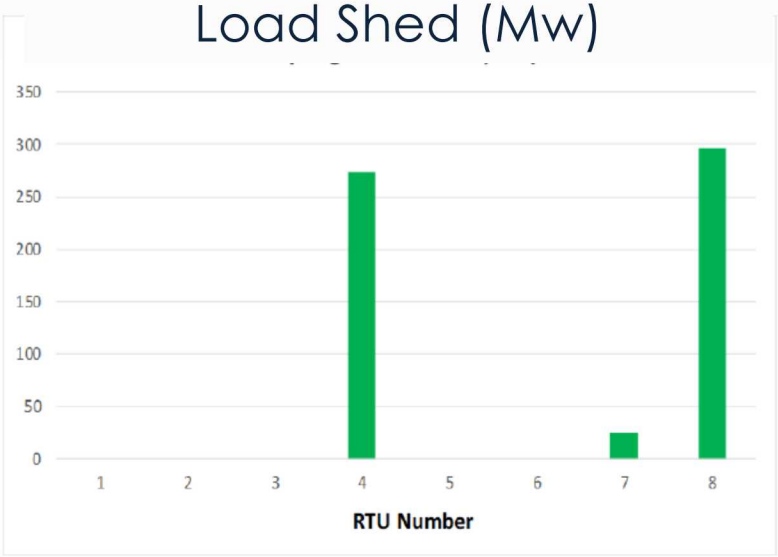
Simulated Grid Case Study: March EAB



Attack RTUs



- RTU Layout:
  - 1 RTU per substation (8 total)
  - Small subset of substations
  - Nominal operations:
    - 320.81 MW Load



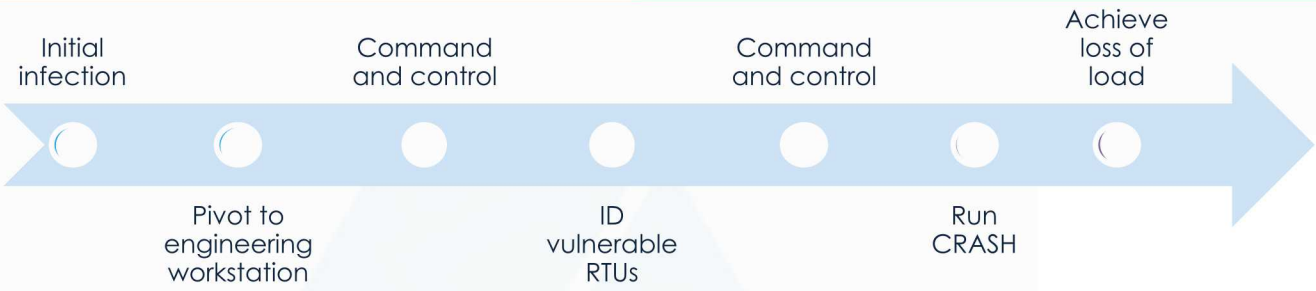
*Derived from synthetic data that does not represent actual grid: https://electricgrids.engr.tamu.edu/electric-grid-test-cases/activsg2000/

Simulated Grid Case Study: March EAB

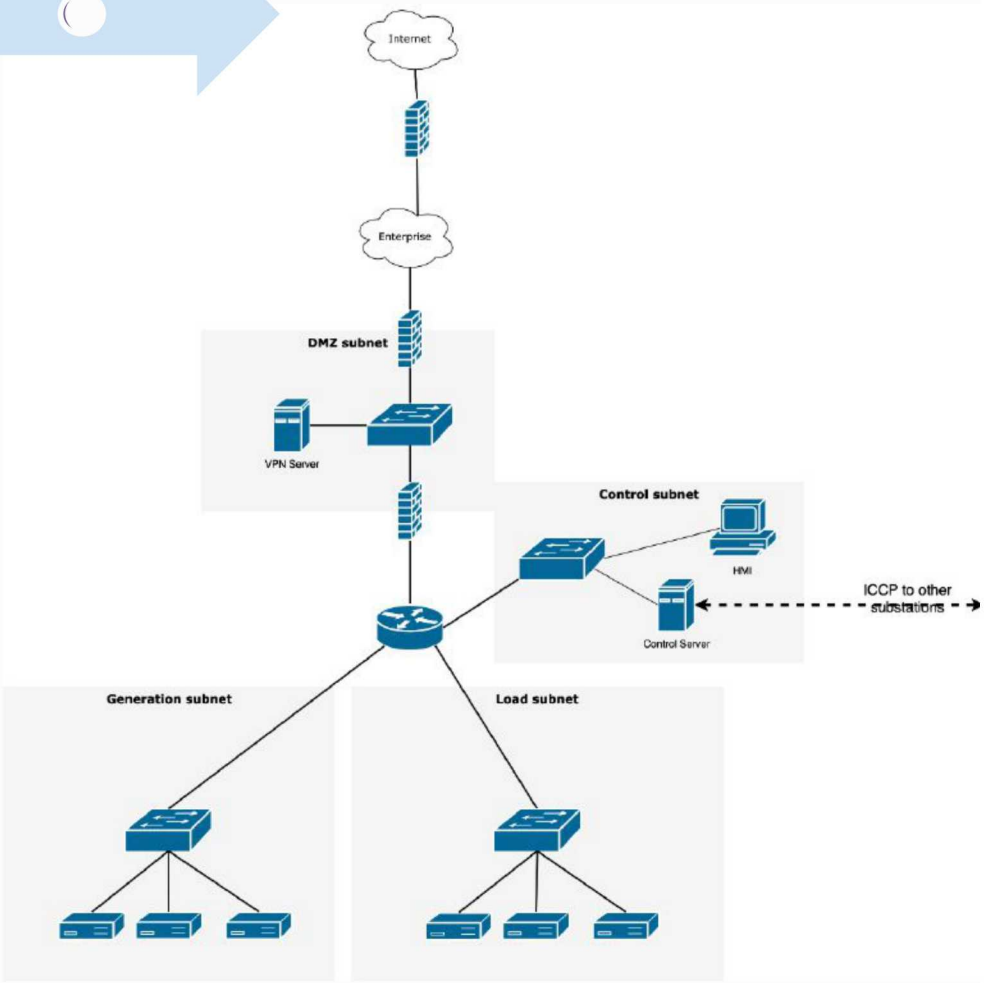
- Key Simplifications
  - Analyzed 1 step of kill chain (action on objective)
  - Considered relatively small system
    - Focused on control network
    - 8 RTUs total
  - Assumed attacker has prior knowledge of RTUs
  - Considered "a lot of grid" and "limited cyber"
    - Single metric of interest: load shed

This presentation describes how we have expanded upon our initial case study to consider additional complexities.

Updated Case Study: Multi-Step Kill Chain

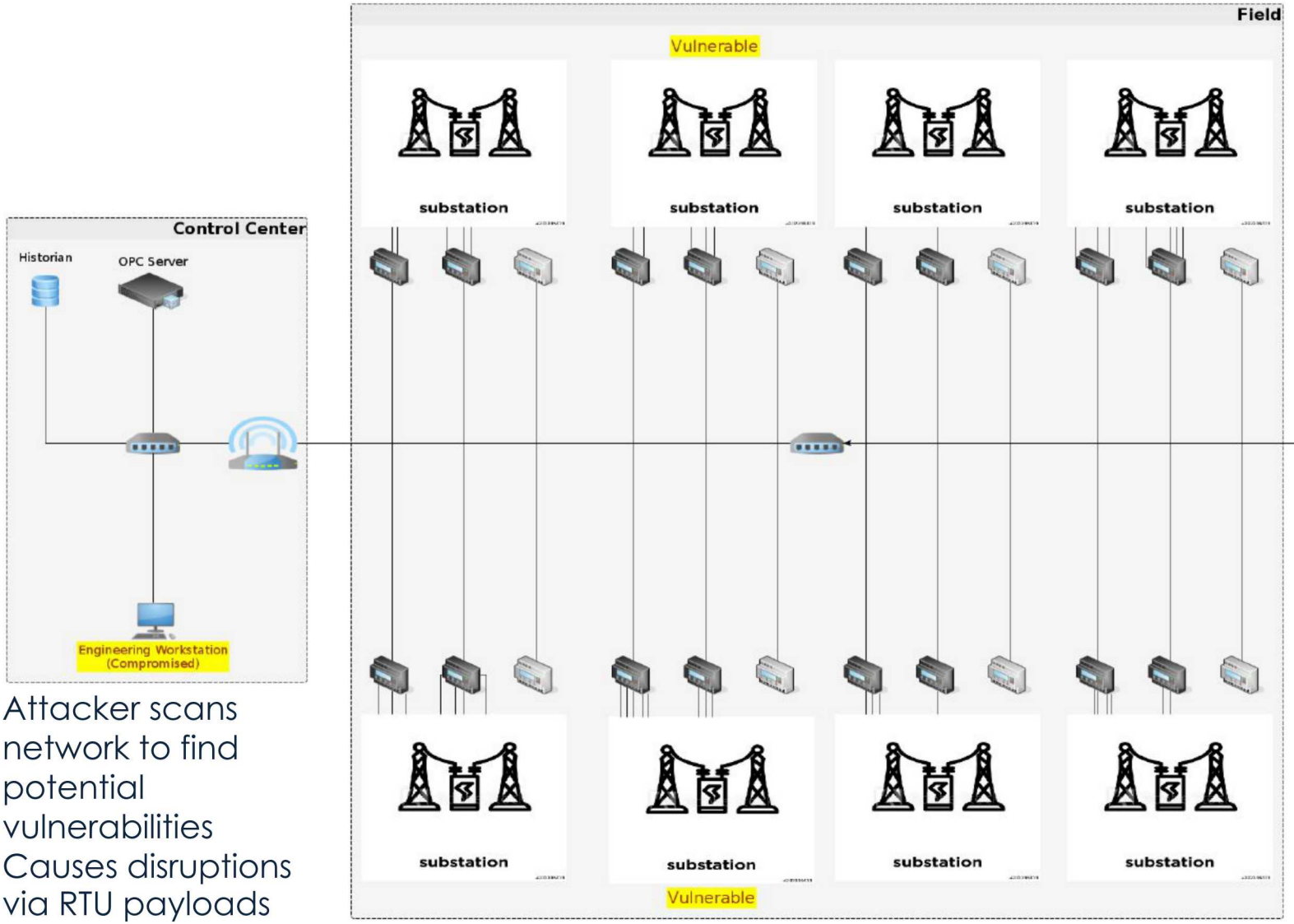


1. Start
2. Deliver email
3. Follow link
4. Execute
5. Obtain IP of engineering workstation
6. Command/control
7. Pivot to engineering workstation
8. Scan for RTUs
9. Ready for attack



Scenario – Cyber
Notional SCADA/ICS Network

8 substations, 24 remote terminal units (RTUs)



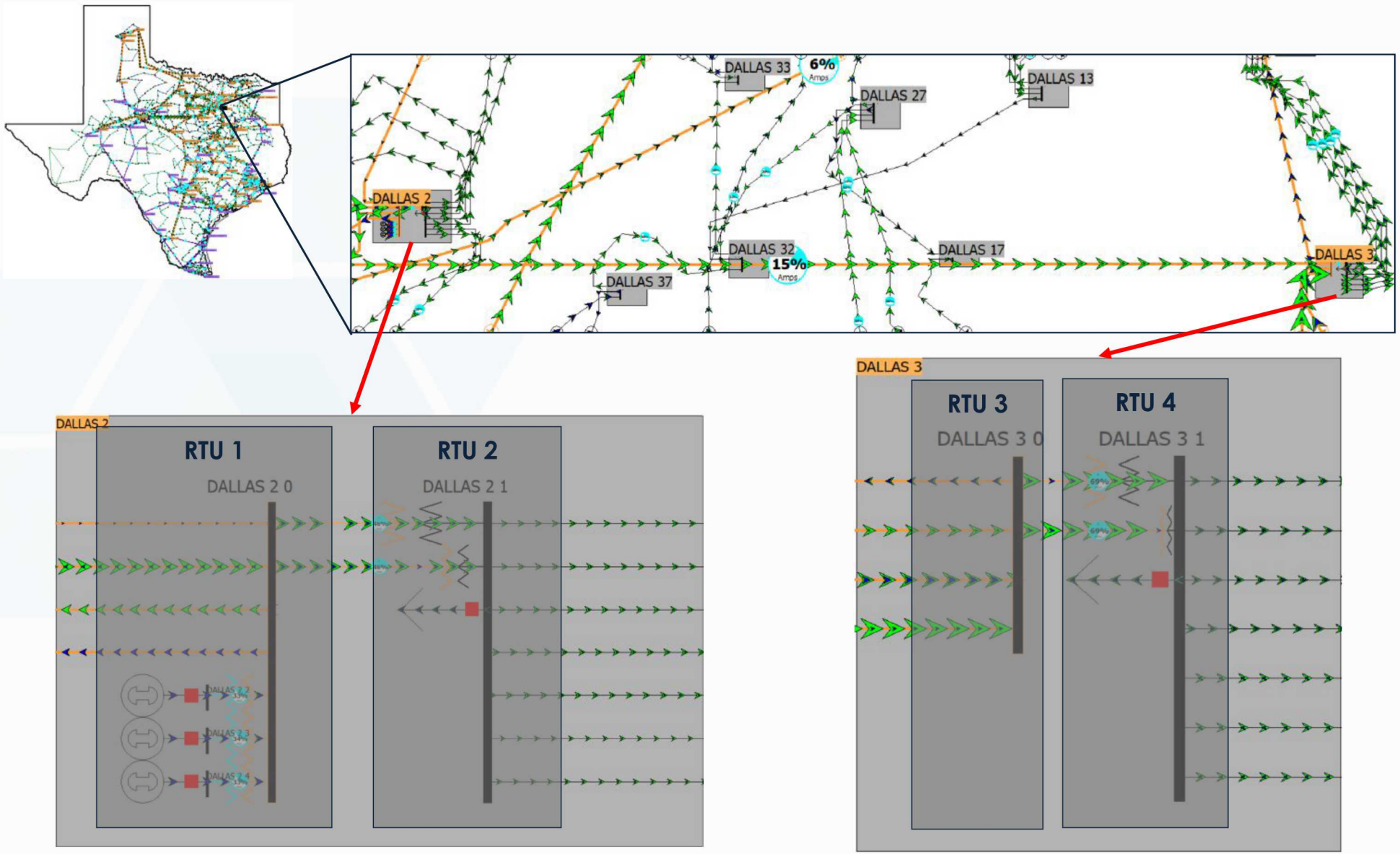
- Attacker scans network to find potential vulnerabilities
- Causes disruptions via RTU payloads

Defender monitors network traffic to detect attacks

Vulnerable RTUs not firewalled for maintenance

Scenario – Physical

2000 Bus Synthetic Model of Texas Power Grid



*Derived from synthetic data that does not represent actual grid: https://electricgrids.engr.tamu.edu/electric-grid-test-cases/activsg2000/

Simulated Grid Case Study: Extensions

	March	October
System Analyzed	Only control network	Enterprise + control networks
Steps in Kill Chain	Single step	Multiple steps
How much cyber and how much grid?	Lots grid, little cyber	Added a lot more cyber
Attacker Knowledge	Knows all vulnerable RTUs	Has to find vulnerable RTUs
Metrics of Interest	Load shed	Load shed + many cyber-focused metrics
Scale	8 RTUs	24 RTUs (and just started on 240 RTUs)

Many of the following presentations will include models, results, analysis, and capability development for portions of this case study.

LDRD

Laboratory Directed Research and Development

SECURE Predictive Cyber Emulation (Emulytics) Task

Tom Tarman

Team members:

- Jerry Cruz
- Sasha Outkin
- Christian Reedy
- Tom Tarman
- Vince Urias
- Eric Vugrin

U.S. DEPARTMENT OF ENERGY

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.

March EAB feedback items covered in this talk

CA Community awareness

- "Future presentations should explicitly acknowledge related research and articulate how SECURE is going beyond it. "

RI Risks identification

- "explicitly identify risks associated with the project and develop strategies to mitigate them "

TC Threat characterization

- "it was not clear to the board how the threat characterization work contributed to the overall goals of the project "

ST SECUREtk

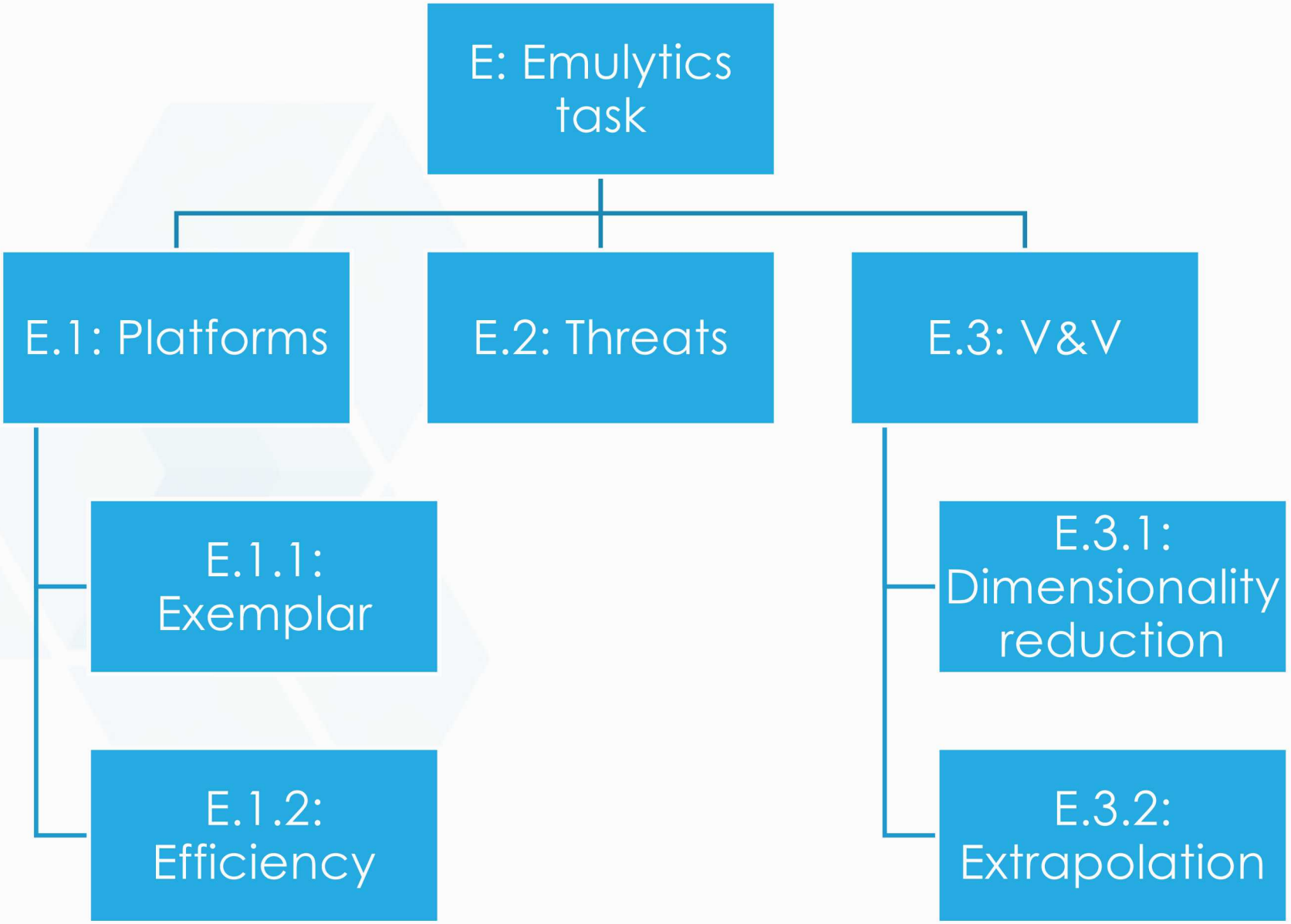
DE Domain expertise

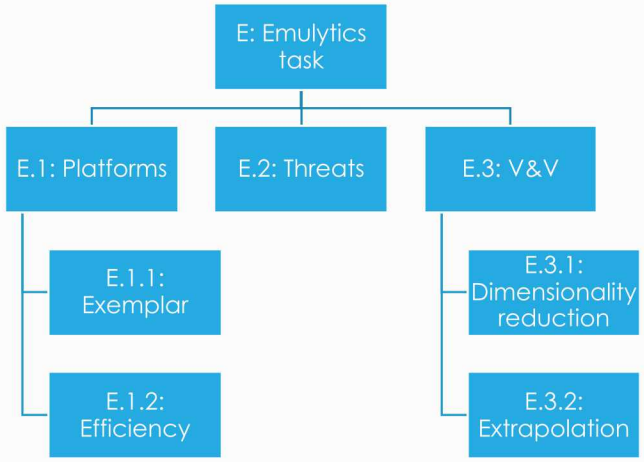
- "given that the exemplar is the power grid, the EAB did not see sufficient evidence that the team has the required domain expertise to create realistic scenarios that will answer meaningful questions "

Predictive Cyber Emulation - Outline

- Task overviews and accomplishments (Tom)
  - E.1: Emulytics platform and modeling
    - Demo overview
  - E.2: Modeling uncertain threats
  - E.3: Model confidence/V&V
- FY20 plans
- Mathematical modeling and validation for network scanning vs. intrusion detection (Eric)

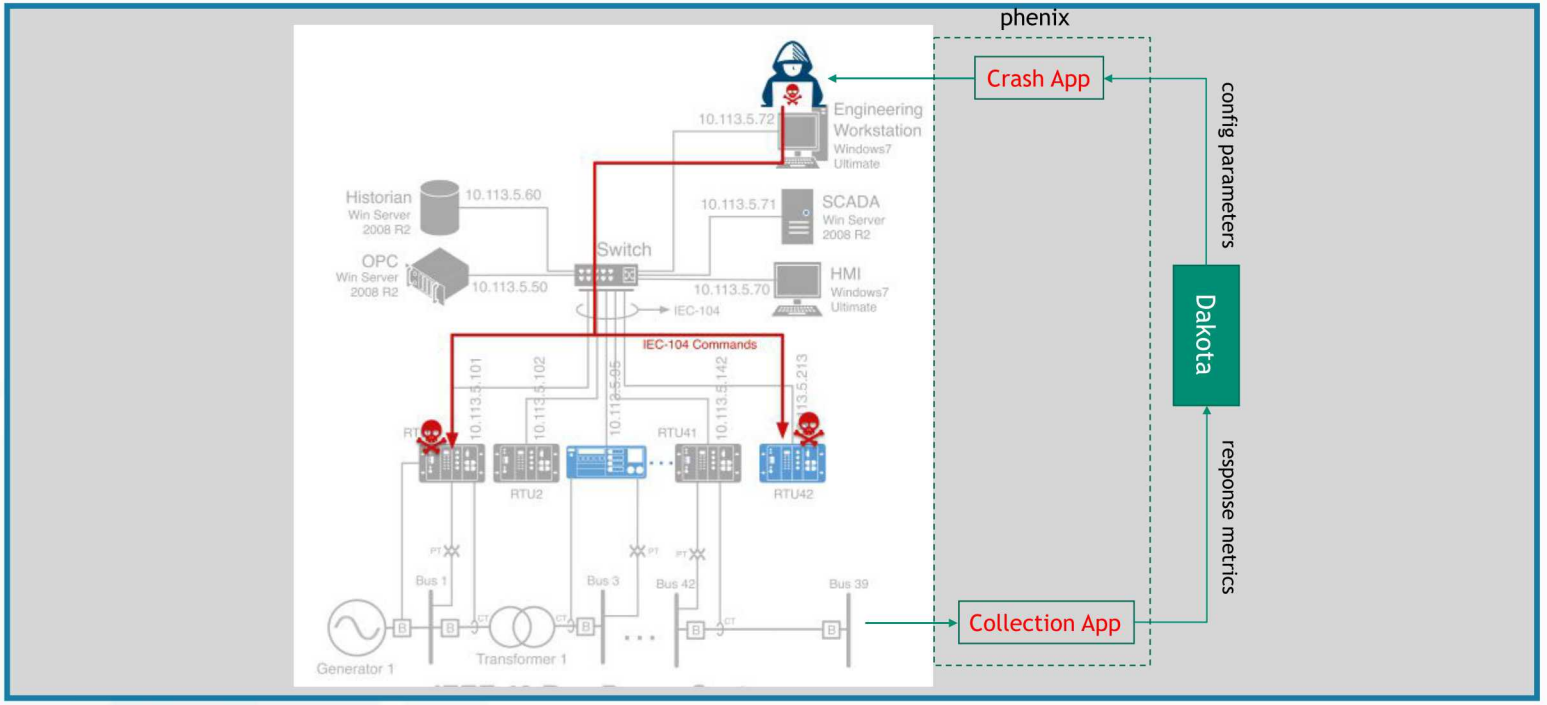
Emulytics task organization





E.1: Emulytics platform and modeling

Research Task E.1: Emulytics platform
Task overview

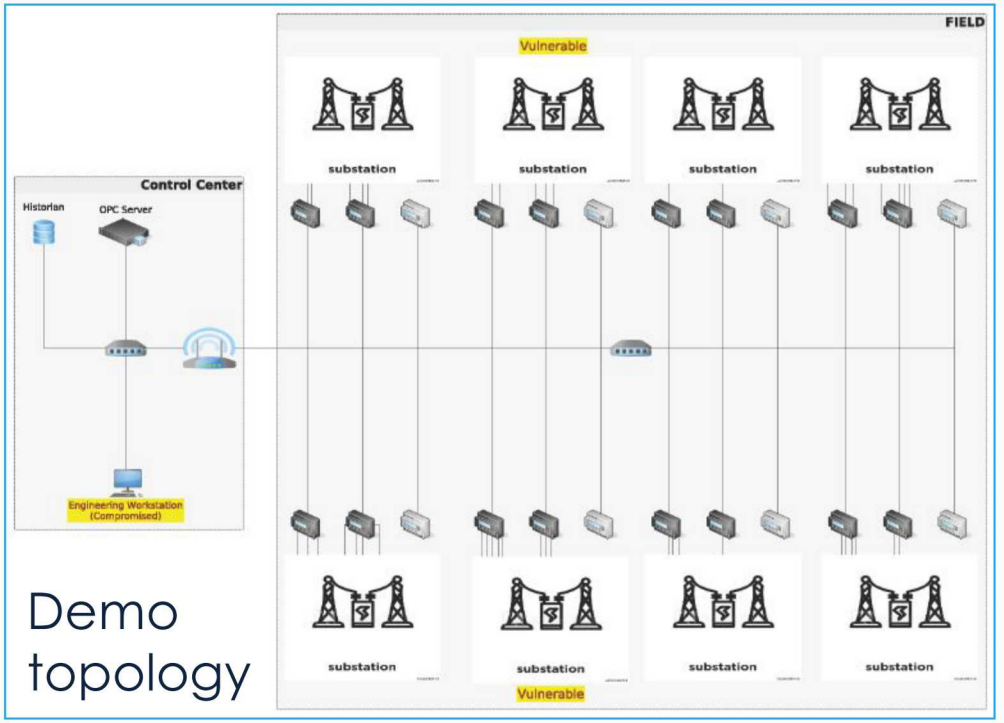


Question: Are there engineering hurdles associated with automated design of experiments and computational efficiency at scale?

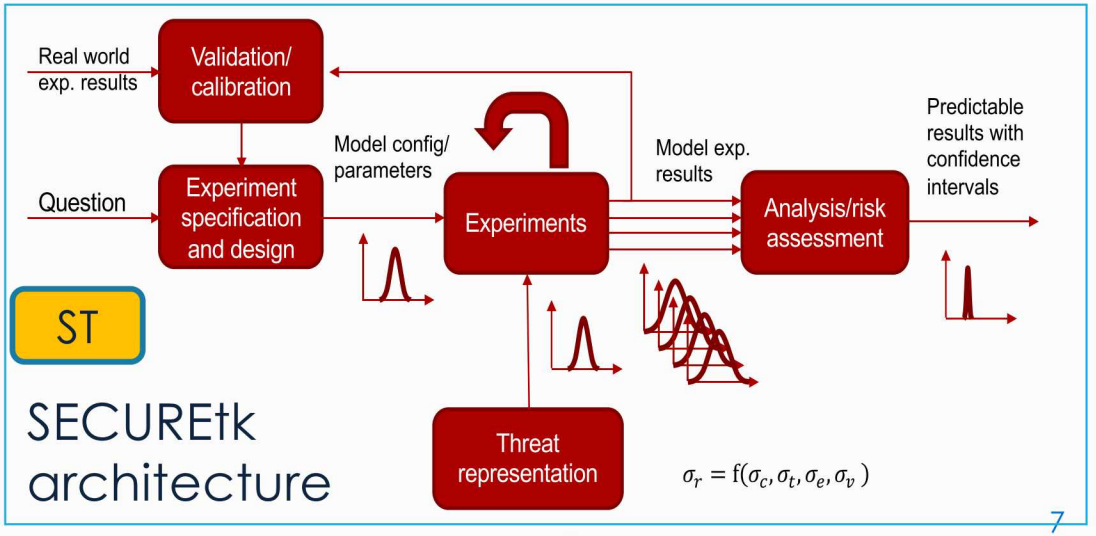
- E.1.1: Develop exemplar questions and models
  - TAMU collaboration will provide tools for topology generation
- E.1.2: Cyber-experimental efficiency and usability
  - Interfaces to allow external control (e.g. Dakota) over parameters and execution
  - Mathematical modeling
  - Experiment platform – (experimental variability, efficiency)
  - SECUREtk (SCORCH)

E.1: Emulytics platform Accomplishments since March

- Defined experimental scenarios
  - Exemplar
  - Optimization and UQ scenarios
- Created and validated mathematical model of network scanning/IDS
- Publications
  - UTSA invited talk
  - CSET CA
  - INFORMS
- External engagements
  - UTSA
  - USC/ISI
  - GA Tech
  - ISO/NE DE
  - TAMU DE



Demo topology



SECUREtk architecture

E.1.1: Exemplar - what is the expected loss of load that results from user receiving a malicious email?

	Initial infection and pivot	Scanning	Command and control	Action on objective
Action	User opens malicious email	Malware scans for RTUs	Malware uploads map and maintains channel for control	Crash override
Questions	How quickly can the threat successfully pivot to the control center network?	How quickly can IDS detect the scan? What fraction of RTUs is detected?	How quickly can IDS detect persistent comms?	What is the loss of load?
Team questions	GPLADD: What is the pivot probability / timing to control center network?	Scanning: What are the scanning time/detection tradeoffs?	Optimization: What is the optimal sensor placement to detect C2?	Optimization: What is the optimal selection of RTUs?
Uncertainties	Location of initial infection	Identified RTUs	Background traffic mix	Timing of attack
Emulytics/multi-fidelity	Malicious emails	Scanners	C2 channels	Crash

Alice the Architect, Grid resilience planning

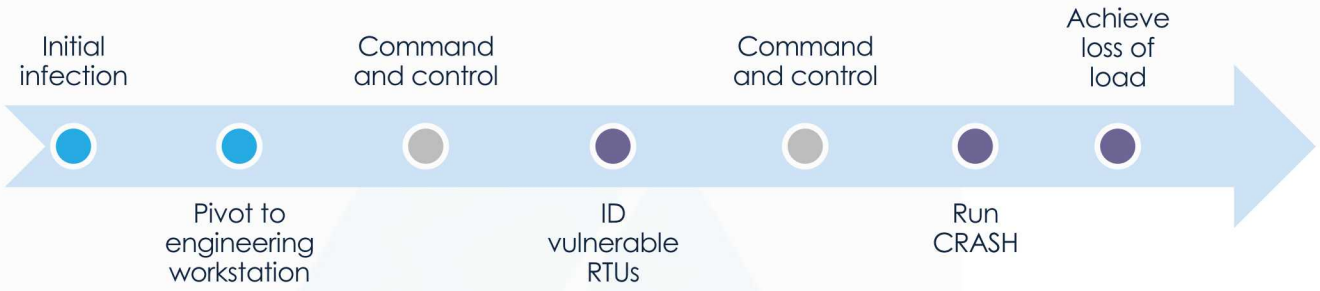
"How do we take into account cyber vulnerabilities in grid operations?"

"How do we take into account consequences in cyber system design?"

Load loss is only one possible metric

GPLADD: Graph-based Probabilistic Learning Attacker and Dynamic Defender
IDS: Intrusion detection system
RTU: Remote terminal unit

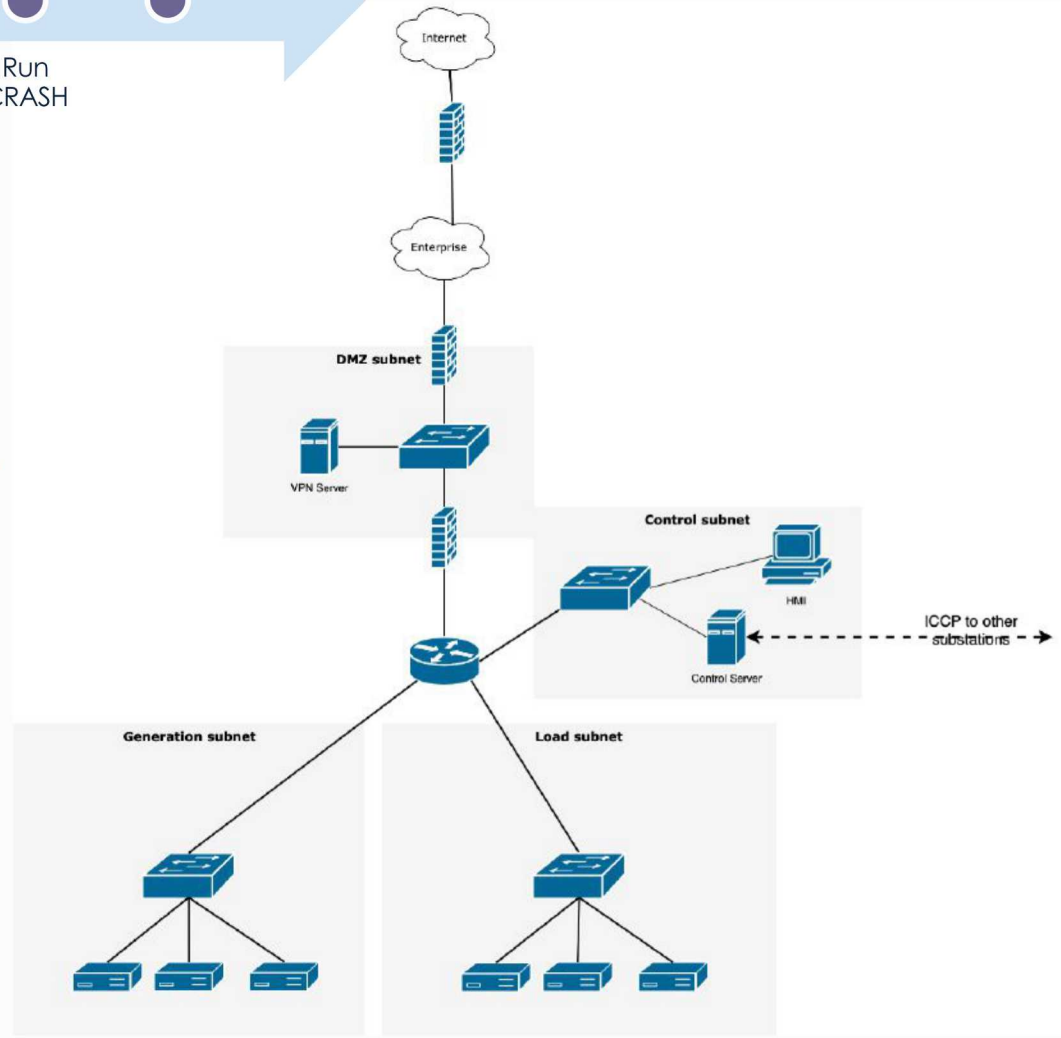
E.1.1: Our approach uses theory and experiment to answer the demo question



Game theory/Markov analysis to assess probability of success and timing

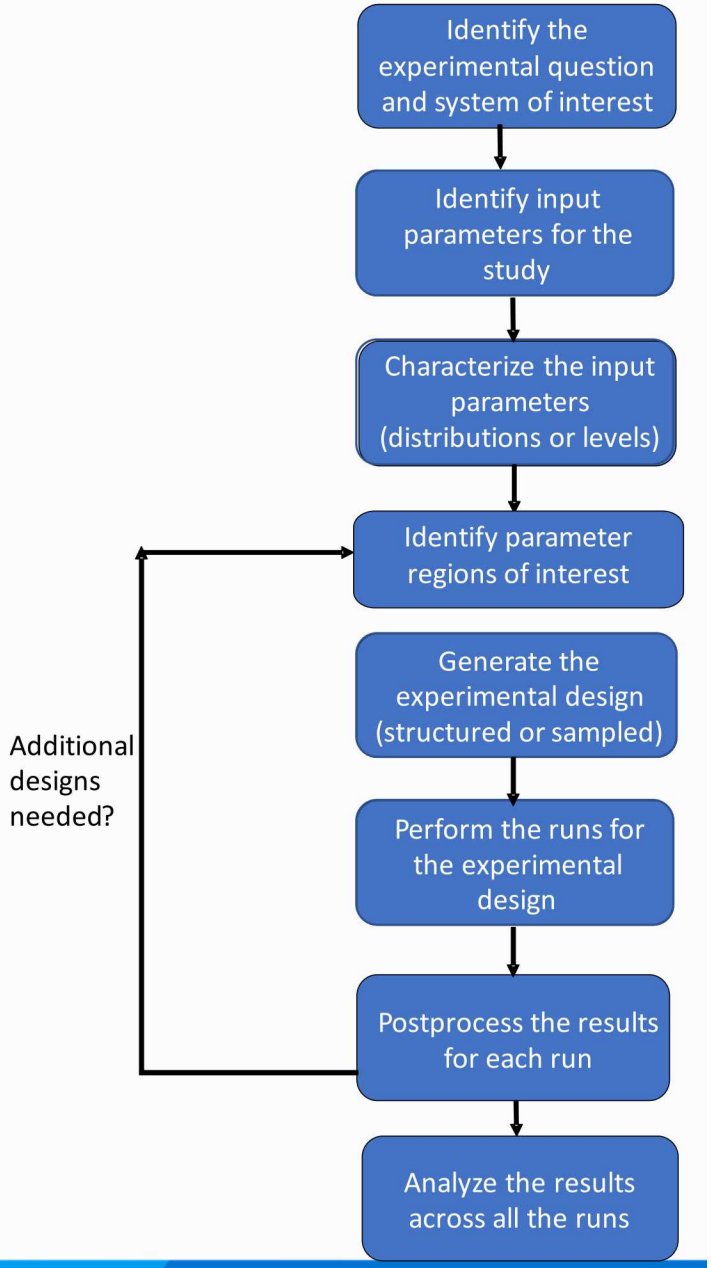
Not assessed at this time

Mathematical and Emulytics modeling to assess number of discovered RTUs and loss of load



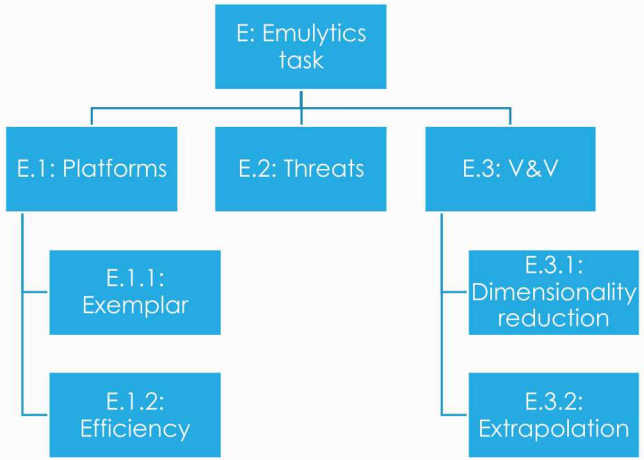
E.1.2: Cyber experimentation efficiency/usability

- Mathematical modeling of network scanning and intrusion detection
  - Topic for today's deep dive
- SCORCH (SCenario ORCHestration)
  - Described in today's deep dive
- Experiment variability
  - Variability due to virtual network interface
    - Inconsistent delays seen in e1000 interface
    - Mitigated by switching to virtio
    - Described in Gianluca's talk
  - Experimental randomness
    - Statistically significant differences between serial and parallel runs
    - Isolated to induced experimental randomness (e.g. packet loss)
    - Described in Laura's talk
- SecureTK
  - Described in Ali's talk this afternoon



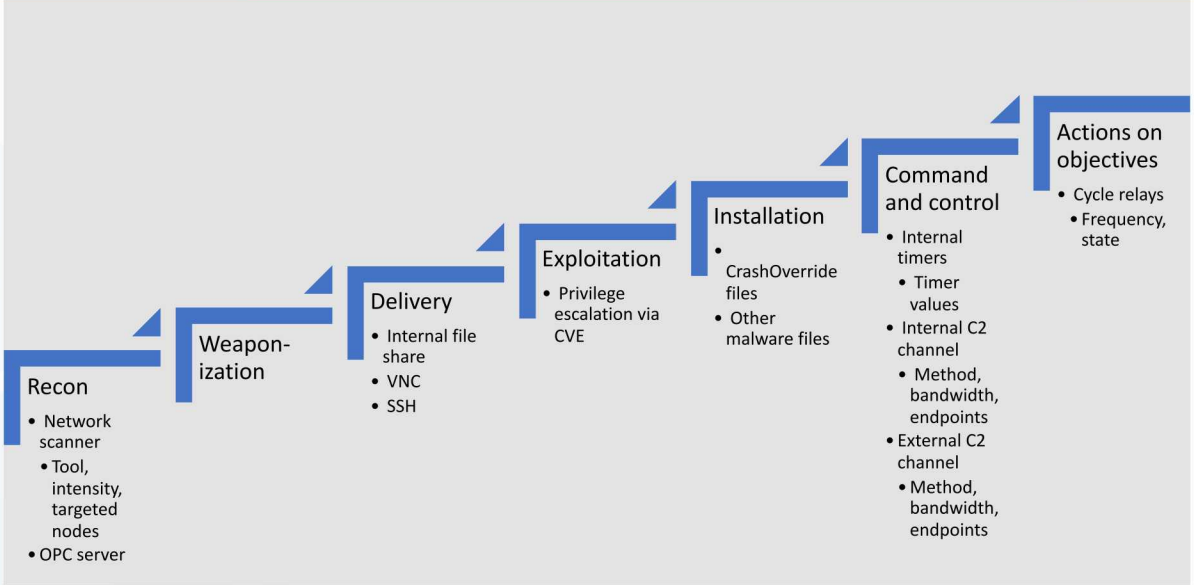
E.1: FY20 plan

- E.1.1: Exemplar demo
  - Propagate Emulytics modeling into the enterprise network
  - Scale up (~100 field devices, control center network, enterprise network)
  - Leverage topologies from TAMU/Kate Davis
  - V&V experiments to support exemplar demo
- E.1.2: Emulytics platform
  - SECUREtk architecture definition
  - Mathematical modeling (e.g. command and control channel)
  - Topology import from Texas A&M models/tools
  - Background traffic
- Publications
  - FY20 - Network scanning/intrusion detection
  - FY20 - Experimental workflow, SECUREtk (or components), case study



E.2: Modeling uncertain threats

Research task E.2: Modeling uncertain threats
Task overview



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.

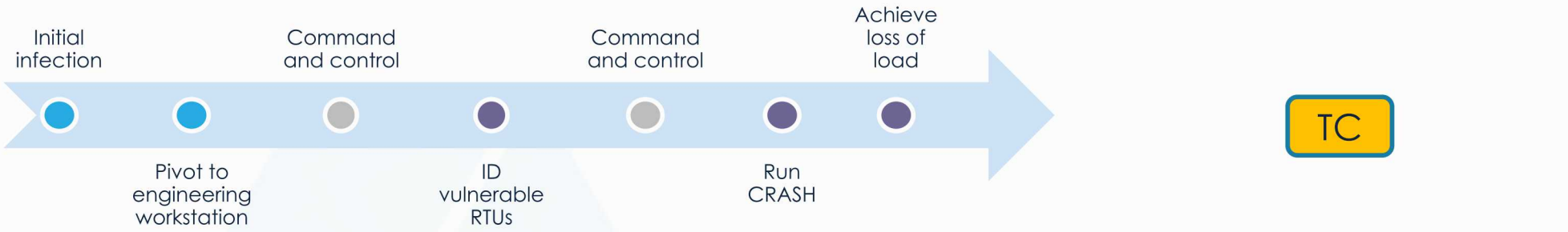
Question: What are the research hurdles associated with modeling sophisticated (and often unknown) threats with uncertainty?

- Specific threats evolve, so adopt frameworks that can be updated as threats change
  - E.g. Lockheed Martin Cyber Kill Chain
  - Game theoretic framework - 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 (initially developed for PRESTIGE hardware trust LDRD)

E.2: Modeling uncertain threats
Accomplishments since March

- Developed Markov threat model for enterprise portion of demo scenario
- Publications (primarily under PRESTIGE LDRD, leveraged by SECURE)
  - Alexander V. Outkin, et.al. GPLADD: Quantifying Trust in Government and Commercial Systems: A Game-Theoretic Approach. ACM Trans. Priv. Secur. 22, 3, Article 18 (June 2019)
  - Yu-Cheng Chen, Dustin Campbell, Vincent Mooney, Santiago Grijalva, Brandon K. Eames, Alexander V. Outkin, Eric D. Vugrin. 2019. "Power Grid Bad Data Injection Attack Modeling in PRESTIGE". Proceedings of 2019 Government Micrcircuit Applications & Critical Technology Conference (GOMACTech)
  - Cynthia Phillips, Alexander Outkin. 2018 "Probabilistic-Learning Attacker, Dynamic Defender: A Cybersecurity Game of Deterrence and Resource Allocation". Workshop on Competitive Economics of Cybersecurity. Albuquerque, NM. November 16, 2018.
- External engagements
  - GA Tech
- Reviewed Dr. Tamer Başar (UIUC) publications CA
  - Attacker/defender modeling in IDS
  - Advances in sensor data aggregation
  - Would complement time-based Attacker/Defender modeling, e.g. to support attack progression inference

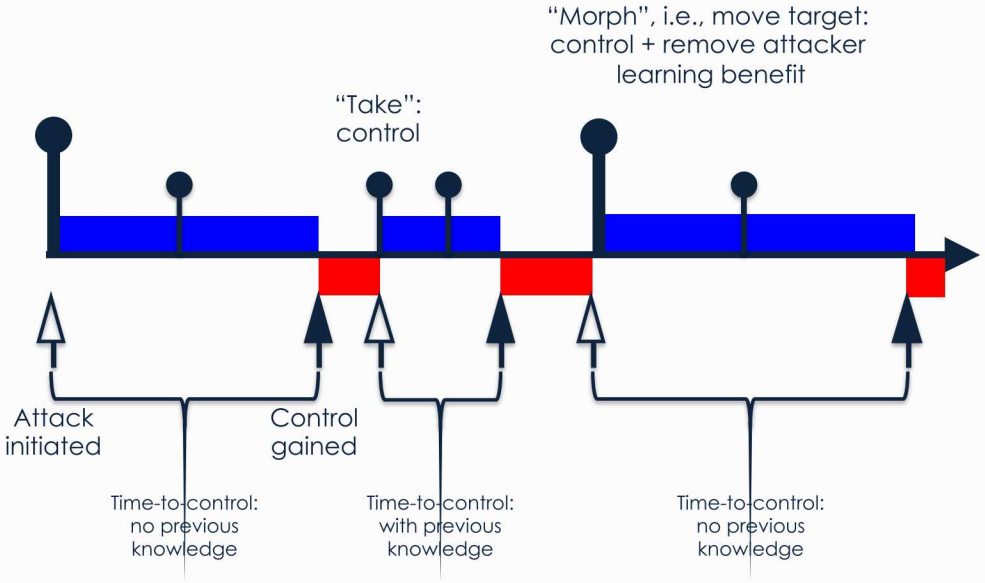
E.2: Why pursue game-theoretic threat modeling?



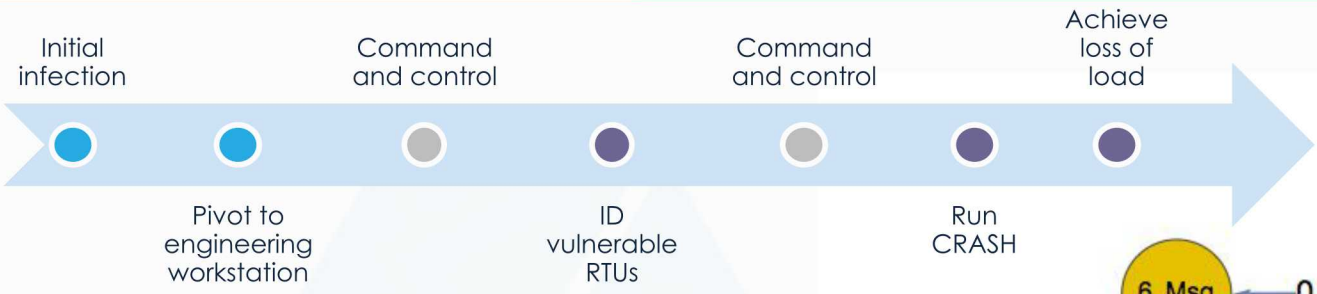
Captain Howard, DoD, high-consequence systems

"Can we credibility assess system performance under various threat scenarios?"

- Game theoretic modeling (GPLADD)
  - End-to-end threat framework
  - Leverages results from in-scope activities (emulation) and out-of-scope data from literature (human factors)
  - Many attacker-defender moves (e.g. moving target defense)
  - Attack/defense evolution over time
  - "First look" at sensitivities that require high fidelity investigation
- Optimization
  - Exact solutions for well-formulated and parameterized problems
  - Identify worst-case threat



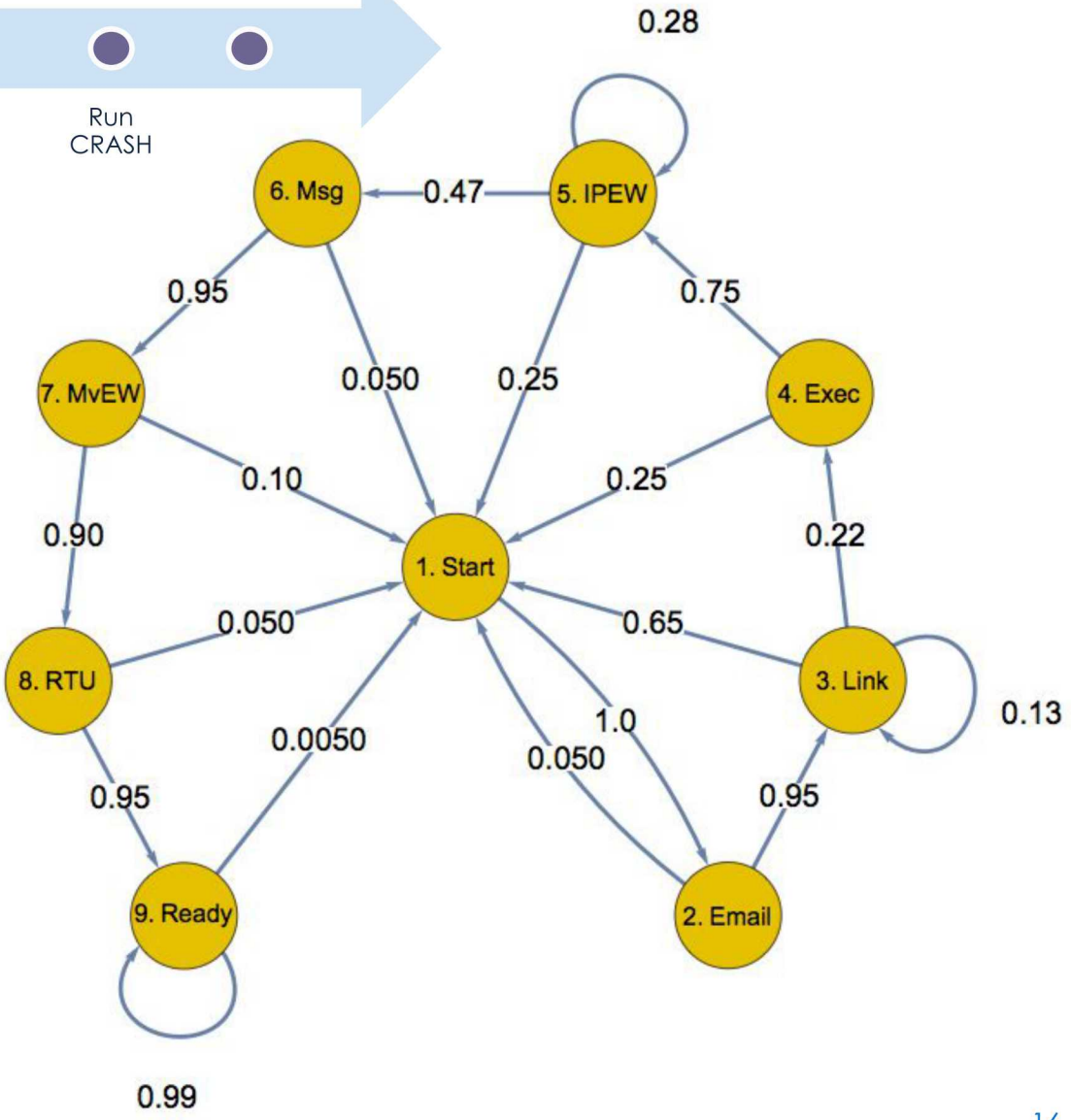
E.2: Markov model is a framework for reasoning about end-to-end threat chain



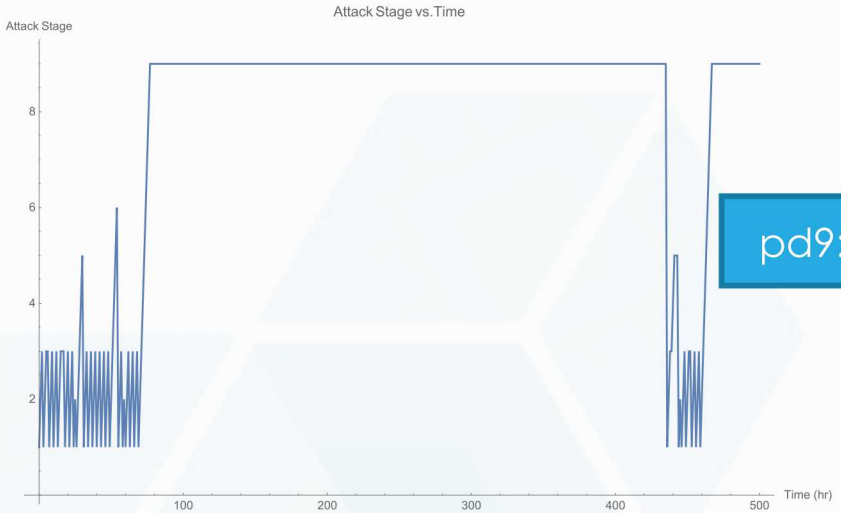
1. Start
2. Deliver email
3. Follow link
4. Execute
5. Obtain IP of engr wkstn
6. Command/control
7. Pivot to engr wkstn
8. Scan for RTUs
9. Ready for attack

Key:

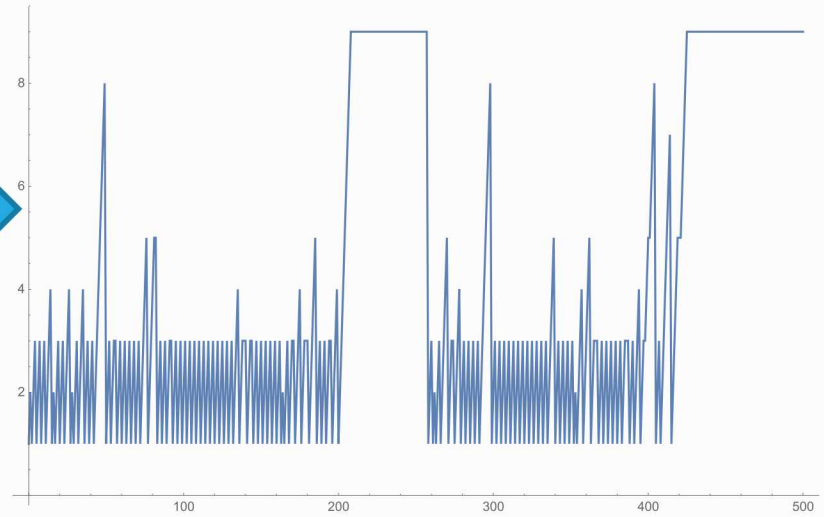
- Informed by literature
- Informed by expert judgment
- Informed by math/Emulytics

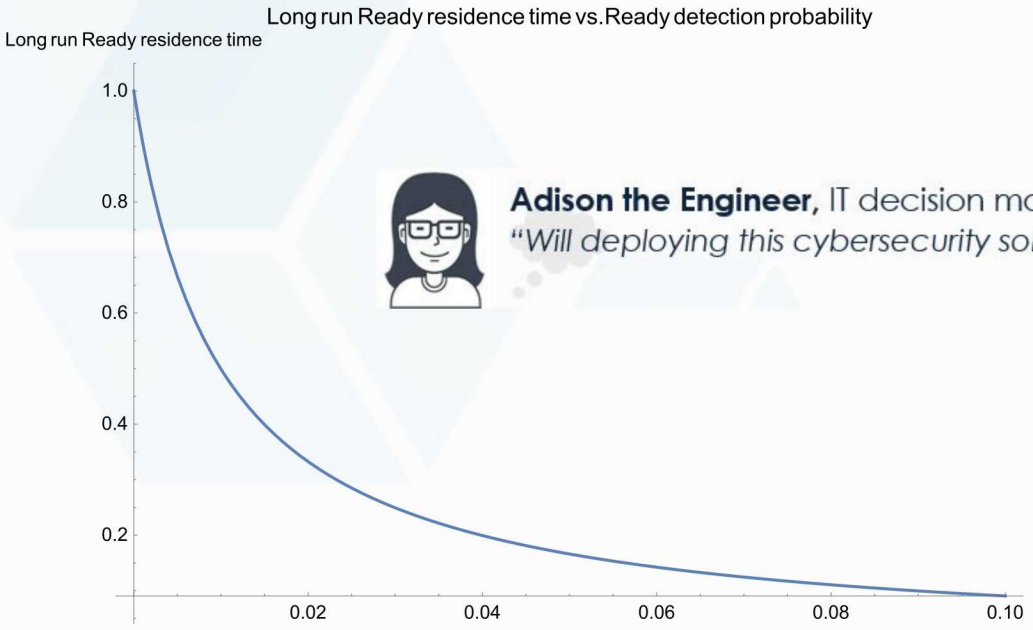


E.2: Preliminary Defender Options Analysis



pd9: 0.005 -> 0.01



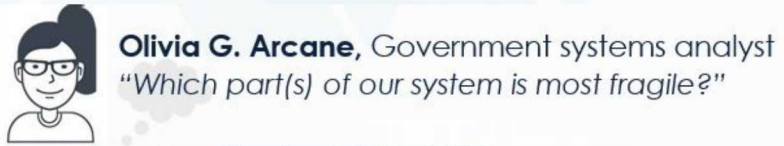


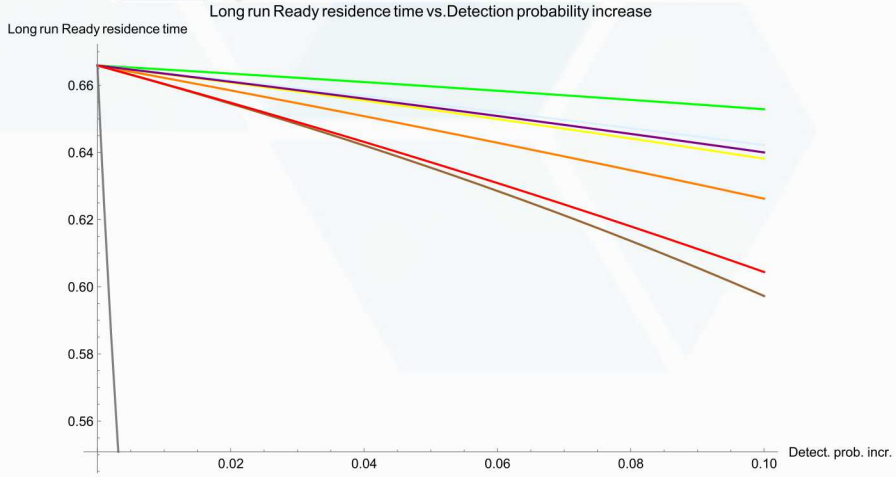
Adison the Engineer, IT decision maker

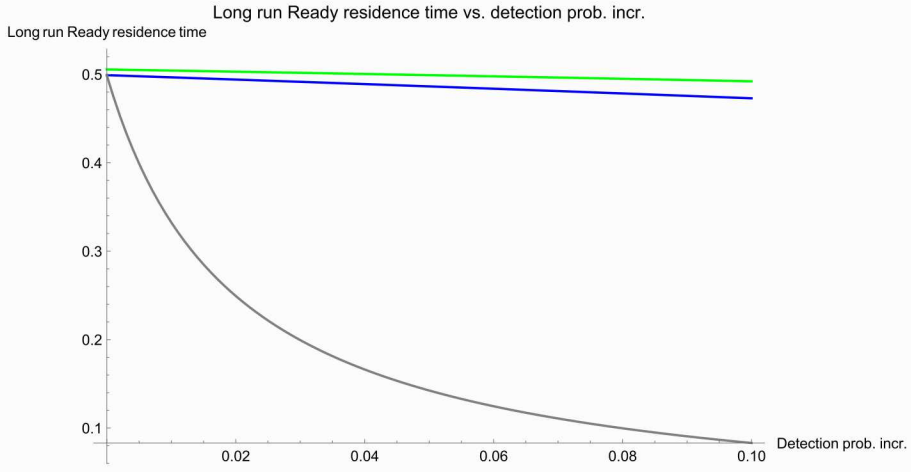
"Will deploying this cybersecurity solution have meaningful impact?"

E.2 Implications to Defender Optimization

- If the cost the defender has to pay per incremental detection improvement on all nodes is the same, then:
  - The defender would get the best return on additional investments in detection improvements on RTUs ("Ready" node 9)
  - Detection improvement on "Ready" node 9 has the least marginal cost
  - The current sensor placement or analytics is locally suboptimal, unless it is a corner solution
  - Next: solve the problem with arbitrary cost functions



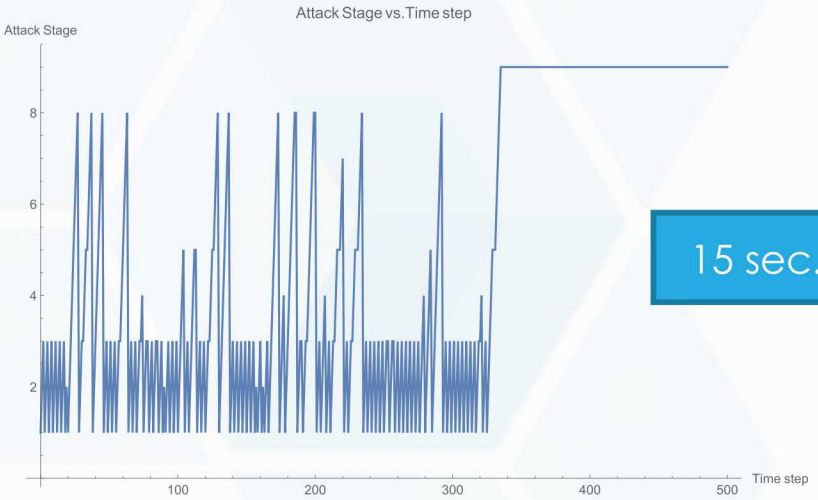




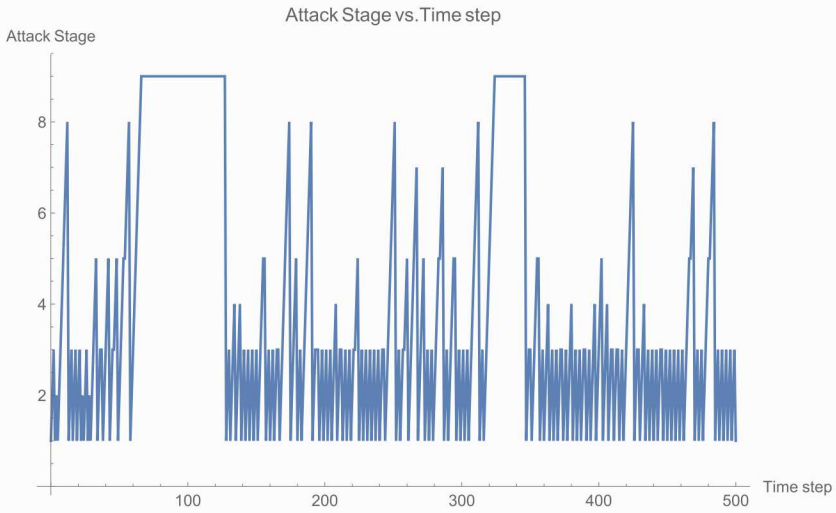
E.2: Comp. of 15 sec vs. 60 sec RTU scans

Scanning strategy (sec)	Long-term attack Ready (5)	Expected load loss, scanning stage (%)	Attack expected load loss (%)	Expected load loss, scanning stage (MW)	Attack expected load loss (MW)
15	0.82	18.41	15.10	84.686	69.44252
60	0.52	32.81	17.18	150.926	79.0480962

End-to-end



15 sec. => 75 sec.



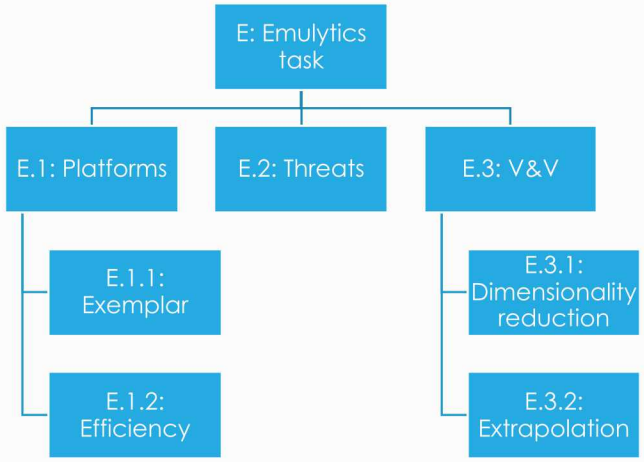
Captain Howard, DoD, high-consequence systems

"Can we credibility assess system performance under various threat scenarios?"

1 timestep = 1 hour

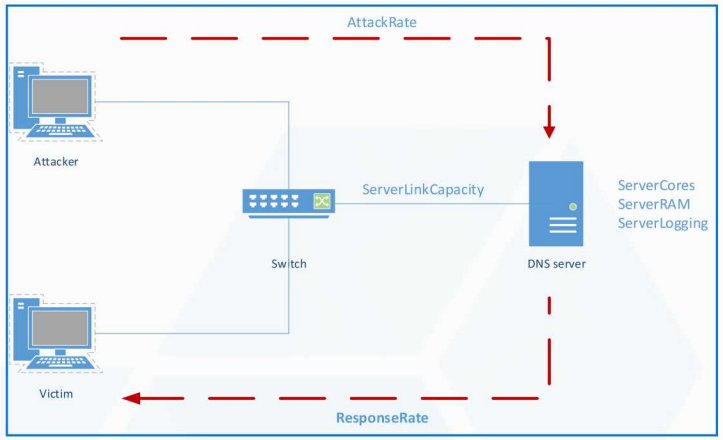
E.2: FY20 plan

- E.2: Modeling uncertain threats
  - Represent a set of attacks
  - Handle uncertainty in threat parameters
  - Add threat models to Emulytics experiment platforms
- Publications
  - FY20 – Markov/GPLADD modeling for power grid threats



E.3: Model confidence/V&V

Research task E.3: Model confidence
Task overview



?

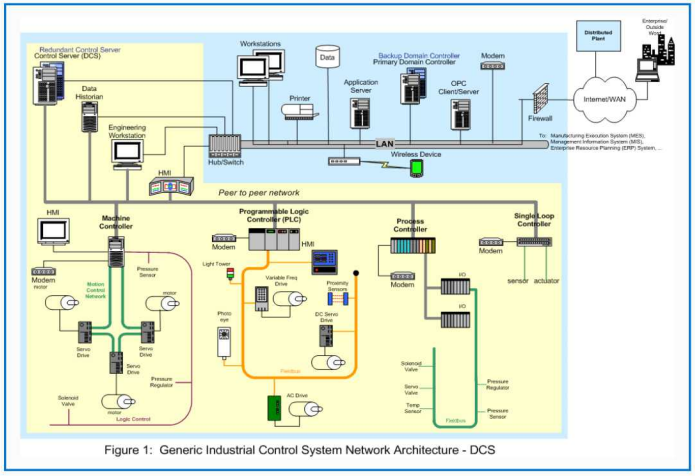


Figure 1: Generic Industrial Control System Network Architecture - DCS

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?

- **E.3.1**: Dimensionality reduction - understand which uncertainties most affect model V&V
  - Kasimir Gabert's research
  - Physical experimentation - collaboration with Kate Davis at Texas A&M
    - RESLab experiments on larger scale ICS systems
    - Funded through LDRD Campus Executive program (Chrisma Jackson, TAMU C.E.)
- **E.3.2**: Extrapolation – understand how V&V experiments extrapolate to V&V statements about larger system

E.3: Model confidence
Accomplishments since March

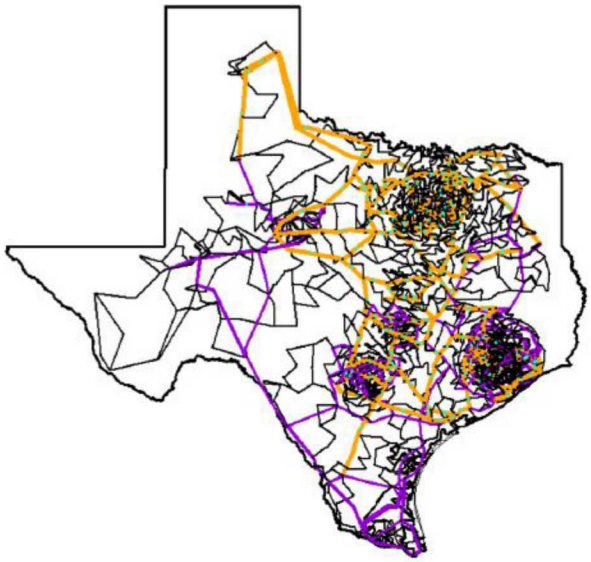
- Worked with Texas A&M to understand their capabilities and identify possible V&V experiments
  - DoS on field devices
  - Protection mechanisms
- Identify sensitive regions using graph dimensionality reduction
  - Kasimir Gabert dissertation at GA Tech
- What do small V&V experiments say about validity of larger systems?
  - UQ team
- Validating mathematical models
- Publications

- External engagements
  - Texas A&M DE

E.3: We are developing a Verification and Validation framework for Emulytics.

- **Overall Question:** "Is our Emulytics model acceptable for a particular application?"
- **Verification:** Is the Emulytics experiment set up so that all VMs operate as if they are running on their own?
  - Is each VM getting all the resources it is requesting? How does host configuration and capacity affect VM behavior?
  - What sanity checks are needed to verify that the VM outcomes are not (or minimally) affected by the run environment?
- **Validation:** Given verification, does the VM produce the same results as a standalone physical node would?
  - What is the impact of behavioral differences in buffer management, network drivers, etc. between virtual environment and physical systems?
  - For which quantities of interests can we make meaningful comparisons using which validation metrics?

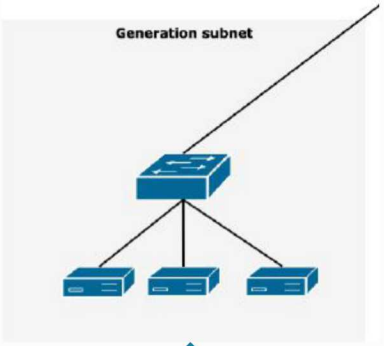
E.3: V&V process



https://electricgrids.engr.tamu.edu/electric-grid-test-cases/activsg2000/

Dimensionality reduction

Extrapolation



Captain Howard, DoD, high-consequence systems
"Can we credibility assess system performance under various threat scenarios?"

Physical V&V

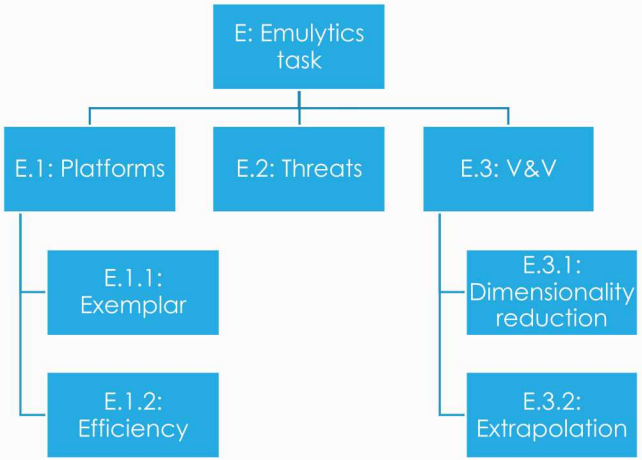
https://upload.wikimedia.org/wikipedia/commons/d/d0/Remote_Terminal_Unit_Modular.jpg

- E.3.1: Kasimir's research – dimensionality reduction
  - Identify repeating subgraphs
  - Summarize large graph with smaller graph
- E.3.1: TAMU/SNL – physical V&V experimentation
  - Protection schemes
  - Response to DoS attack
- E.3.2: UQ team research - extrapolation
  - Small V&V → Large system

*Derived from synthetic data that does not represent actual grid: https://electricgrids.engr.tamu.edu/electric-grid-test-cases/activsg2000/

E.3: Ongoing and FY20 Planned V&V Activities:

- Verification
  - Examples:
    - Ensure that equivalent network paths have similar routing statistics
    - Comparison of statistics from serial and parallel runs
  - Goals:
    - Develop a set of necessary conditions for verified Emulytics
    - Incorporate tools for assessing these conditions in SECUREtk
    - Apply these tools to SECURE exemplars
- Validation:
  - Examples:
    - Compare analytic scanning model to Minimega
    - Compare Bilevel optimization to Powerworld
    - Compare Emulytics models to physical testbeds (TAMU)
  - Goals:
    - Set up physical testbeds
    - Develop appropriate metrics: e.g. QoI distributions or sensitivities
    - Develop approach for validating larger scale systems built from smaller scale validated components: Kasimir's dissertation work
- Publications
  - FY21 - V&V (dimension reduction and extrapolation)


E: Emulytics task
E.1: Platforms
E.2: Threats
E.3: V&V
E.1.1: Exemplar
E.1.2: Efficiency
E.3.1: Dimensionality reduction
E.3.2: Extrapolation

FY20 plans

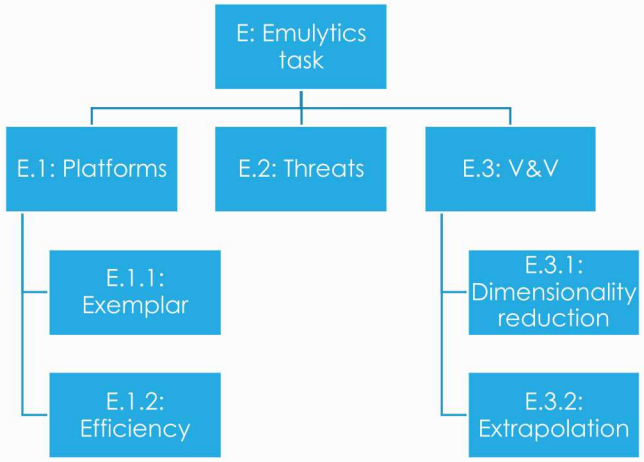
FY20 plan

- E.1.1: Exemplar demo
  - Propagate Emulytics modeling into the enterprise network
  - Scale up (~100 field devices, control center network, enterprise network)
  - Leverage topologies from TAMU/Kate Davis
  - V&V experiments to support exemplar demo
- E.1.2: Emulytics platform
  - SECUREtk architecture definition
  - Mathematical modeling (e.g. command and control channel)
  - Topology import from Texas A&M models/tools
  - Background traffic
- E.2: Modeling uncertain threats
  - Represent a set of attacks
  - Handle uncertainty in threat parameters
  - Add threat models to Emulytics experiment platforms
- E.3: Model confidence
  - V&V experiments with Texas A&M
  - Graph theoretical network dimensionality reduction for V&V
- Publications
  - FY20 - Network scanning/intrusion detection
  - FY20 - Experimental workflow, SECUREtk (or components), case study
  - FY20 – Markov/GPLADD modeling for power grid threats
  - FY21 - V&V (dimension reduction and extrapolation)

Risks

RI

	Risk	Next steps
Exemplar demo	Unrealistic topologies	TAMU collaboration
Emulytics platform	Experimental variation	Understand variability and possible mitigations
Threat uncertainty	Are GPLADD/Markov models valid?	Start integrating threat tools into models
Model confidence	Invalid extrapolation	Understand conditions for valid extrapolation



Scanning – mathematical modeling, emulation, and validation

Eric Vugrin

LDRD

Laboratory Directed Research and Development

SECURE Predictive Cyber Emulation: Example Application to a Scanning & Detection Scenario

Team members:

- Jerry Cruz
- Alexander Outkin
- Christian Reedy
- Tom Tarman
- Vince Urias
- Eric Vugrin

Presenter: Eric Vugrin

U.S. DEPARTMENT OF ENERGY

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.

Preview

- Scenario: investigated use of scanning & detection tools during the reconnaissance phase of an attack on the power grid

- Method:
  - Developed Emulytics model of attack
  - Developed mathematical, "glass box" model of attack

- Results:
  - Validated glass box model against Emulytics experiments
  - Glass box model and enhancements to Emulytics infrastructure have resulted in computational and analytical efficiencies for studying relevant uncertainties

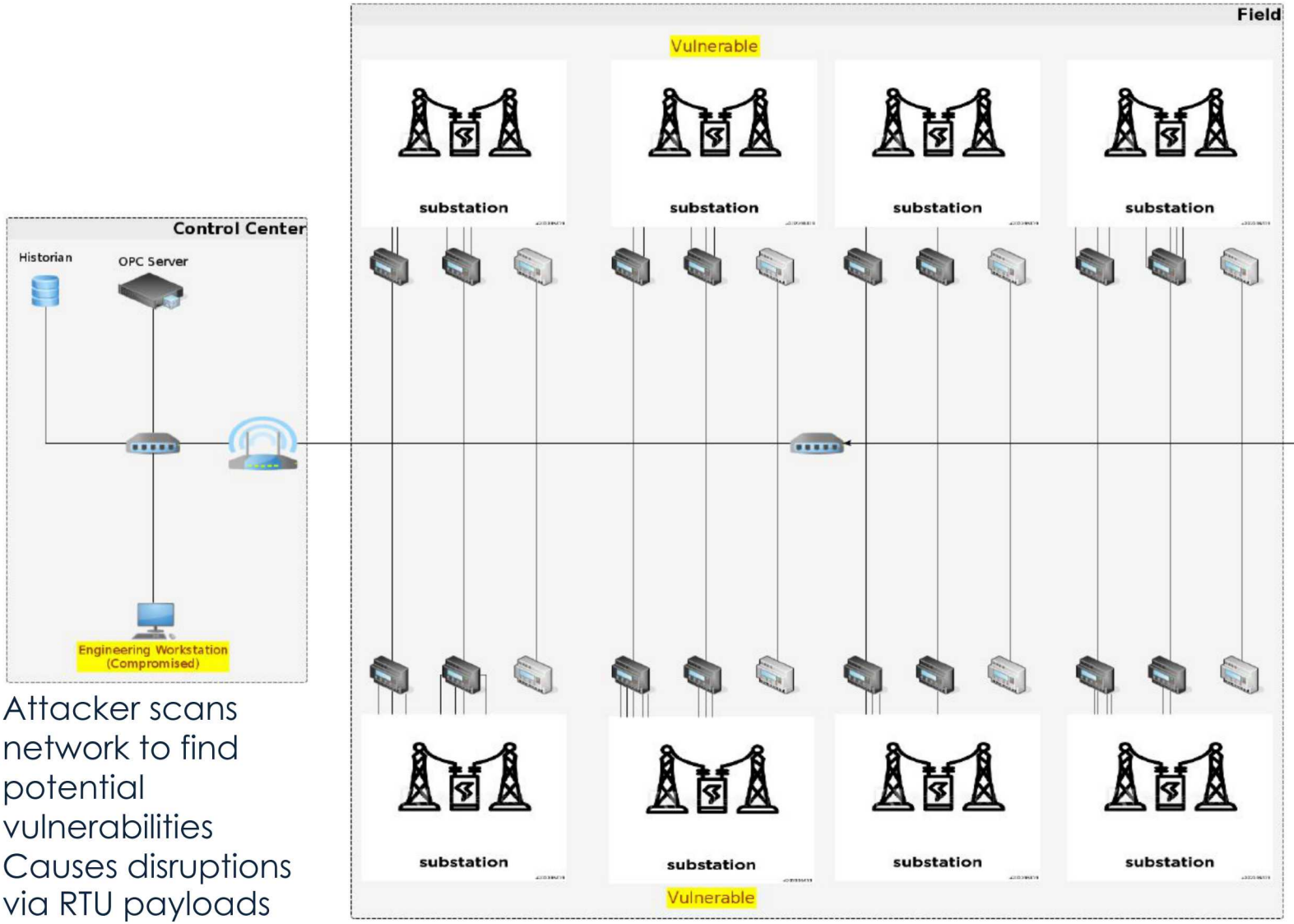
Outline

- Introduce scenario
- Specify analysis and research questions
- Describe methodology
- Provide results
- Discuss insights and future directions

This presentation aims to show some progress towards Emulytics research goals in the context of a specific scenario.

Scenario – Cyber
Notional SCADA/ICS Network

8 substations, 24 remote terminal units (RTUs)



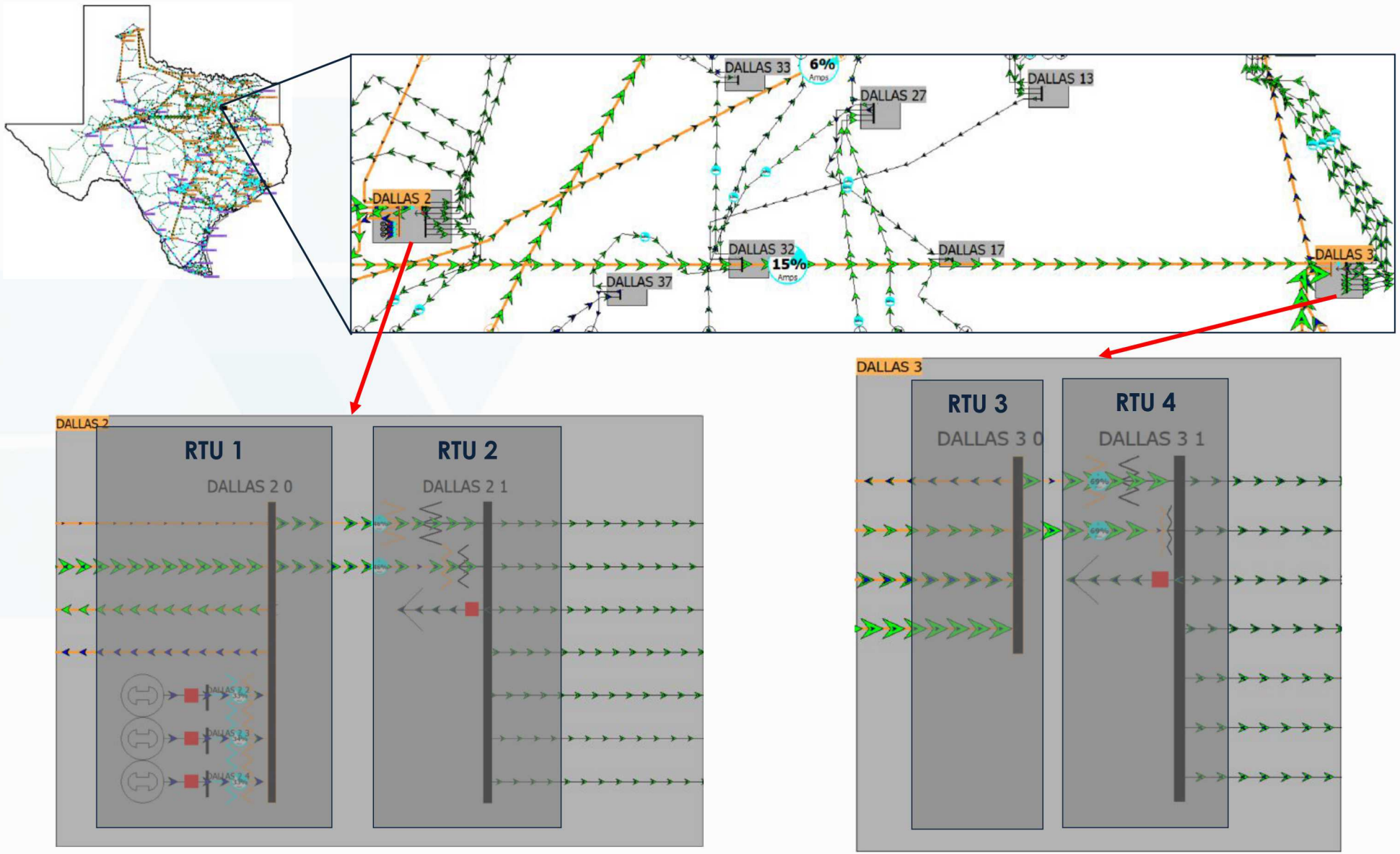
- Attacker scans network to find potential vulnerabilities
- Causes disruptions via RTU payloads

Defender monitors network traffic to detect attacks

Vulnerable RTUs not firewalled for maintenance

Scenario – Physical

2000 Bus Synthetic Model of Texas Power Grid



*Derived from synthetic data that does not represent actual grid: https://electricgrids.engr.tamu.edu/electric-grid-test-cases/activsg2000/

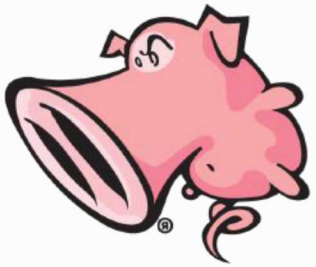
Assumptions

Attacker

- **Goal**:
  - Find vulnerable RTUs quickly & stealthily
  - Cause loss of load
- **Tool**: NMap Network Mapper

Defender

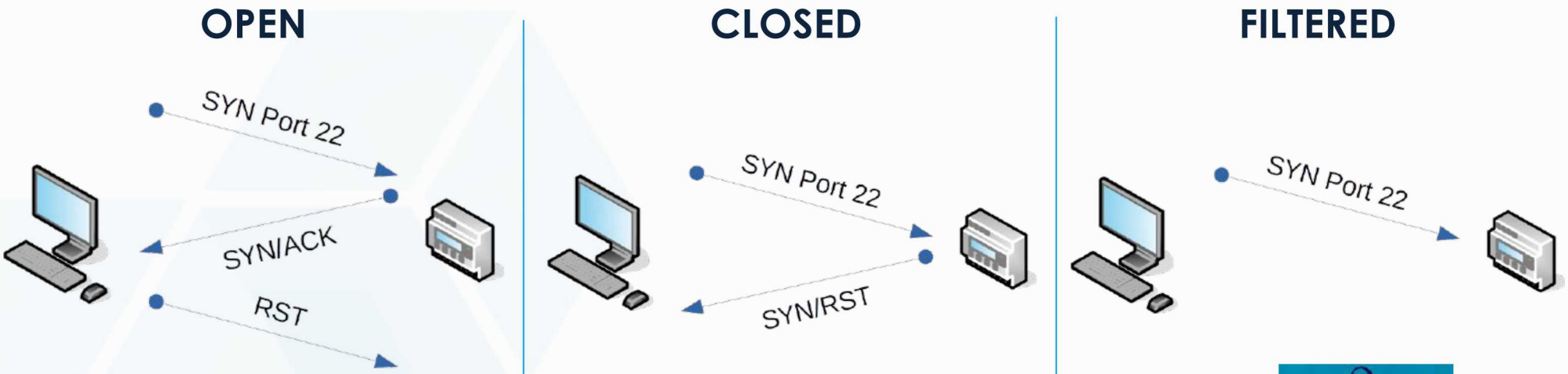
- **Goal**:
  - Detect attack before attacker can exploit vulnerabilities
- **Tool**: Snort



Assumptions: Tools - NMap

TC

- Half-open SYN scan (Bou-Harb et al. (2014))



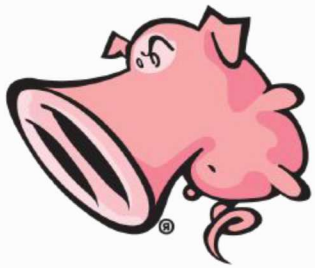
- Key parameters
  - **Host Group Size** – The number of hosts to scan in parallel
  - **Delay** – The delay time between sequential probes
- Assumption: Which hosts are up is known
  - Accomplished via initial ping scan (ICMP echo requests) in the emulation

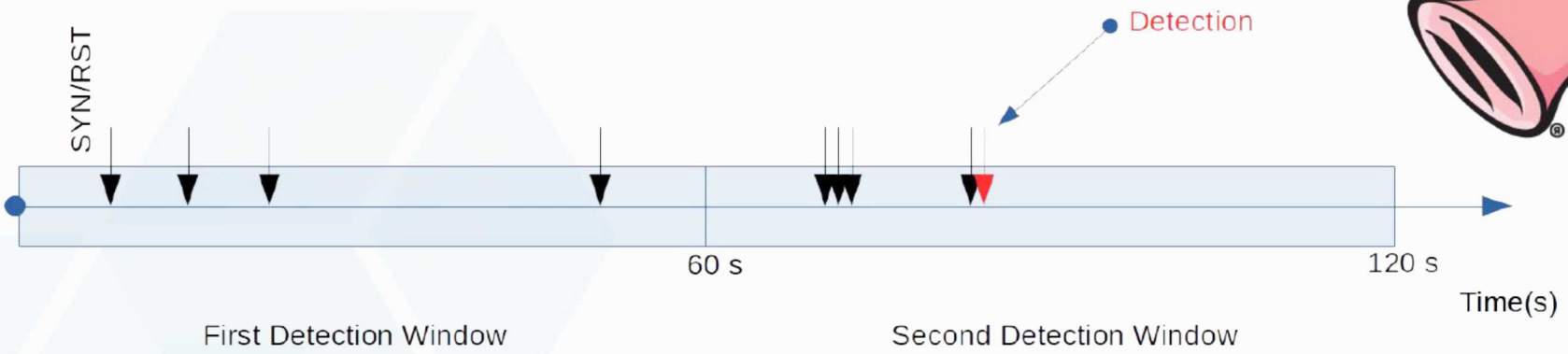
Bou-Harb et al. (2014). "Cyber Scanning: A Comprehensive Survey," IEEE Communications Surveys & Tutorials, 16(3): 1496-1519.

```
nmap -PE -sS -n -p 22 --min-hostgroup 4 --max-hostgroup 4 --scan-delay 10s --min-rtt-timeout .5s --max-rtt-timeout .5s --max-retries 1 --randomize-hosts 10-10.0-0.1-1.1-24
```

Assumptions: Tools - SNORT

- sfportscan module (Roelker et al. 2004)

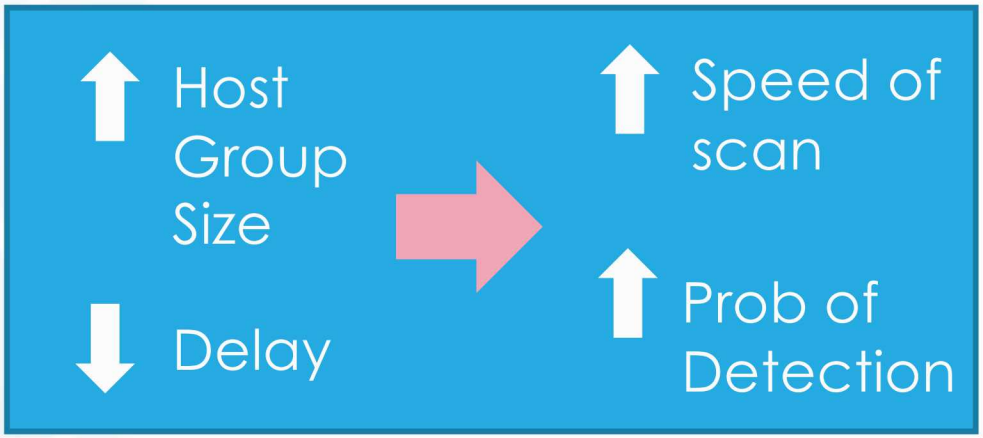




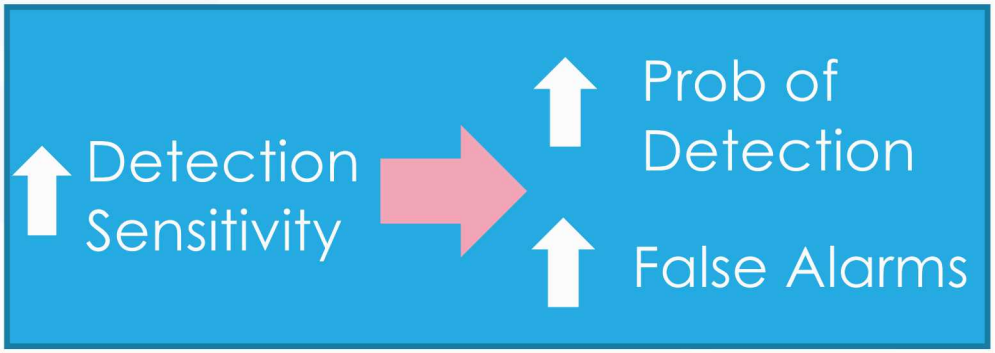
- If snort observes 5 or more TCP SYN/RSTs (during initial 3-way handshake) within a 60 second window, it creates an alert (i.e. detection)
  - An NMap probe to a closed port generates this kind of reset
  - Medium and high sensitivity are similar but with different thresholds and they also count number of new TCP connections
- Assumption: normal traffic does not result in TCP SYN/RSTs

Assumptions: Tools- Effects of Key Parameters

NMap (Attacker)



Snort (Defender)



Questions

Adison the Engineer

- Analysis: for this scenario, can we estimate
  - Rate of vulnerable RTU identification?
  - Probability that the attacker is detected over time?
  - At which point during the scan should the attacker attack RTUs to maximize loss of load?

Wally the Attacker

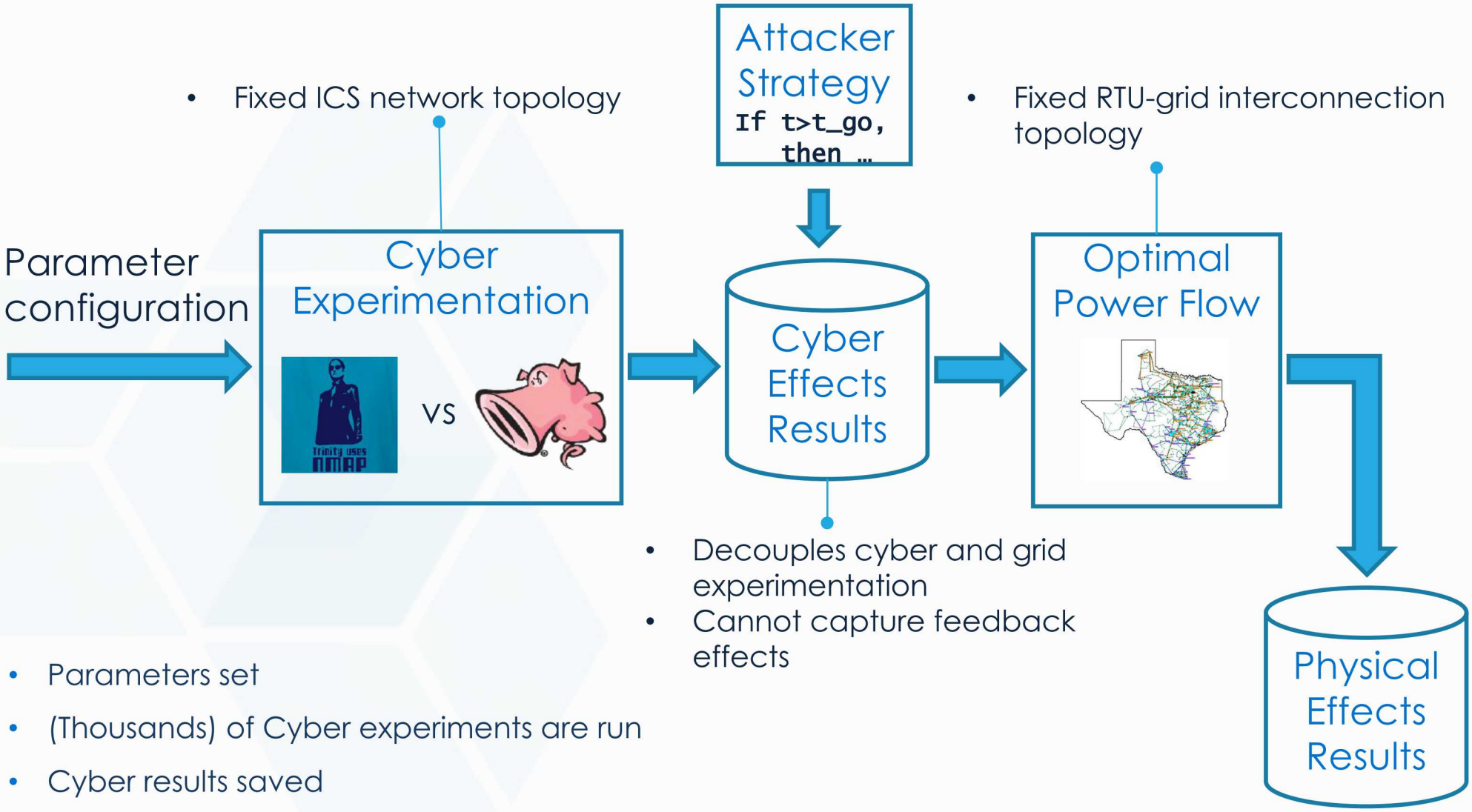
- Validation: can we validate results from emulation experiments through comparison with glass box model estimates? And vice versa?

Dr. Turing the PI

Leon the PM

- Practical consideration: how can we implement experiments in organized, efficient manner to capture potential uncertainties?

Experimental Workflow



- Parameters set
- (Thousands) of Cyber experiments are run
- Cyber results saved
- Cyber results are translated to inputs to optimal power flow tools
- OPF tools generate physical effects results

Uncertainty within the Analysis

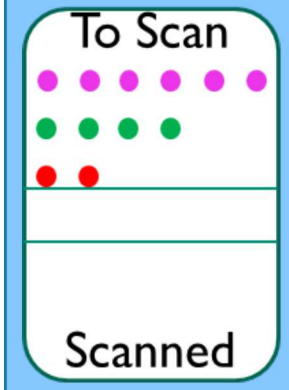
- Sources of uncertainty:
  - Order of scanning RTUs
  - Time out of scanning probes
  - RTUs discovered
- Treatment within Experiments
  - Emulation experiment repeated 1000 times
  - Each experiment run for ~200 seconds
- Outputs
  - Vulnerable RTUs discovered vs time
  - Probability attacker is detected vs time
- Loss of load estimation
  - Attacker strategy specifies when to attack RTUs
  - Post-process cyber effects to determine if RTU attack starts before detection
  - If so, determine which RTUs were identified and use look up table to determine load loss

RTUs Attacked	Load Loss (relative to max=460MW)
1	0.00
2	0.48
3	0.17
4	0.52
1,2	0.48
1,3	0.00*
1,4	0.52
2,3	0.48
2,4	1.00
3,4	0.52
1,2,3	0.48
1,2,4	1.00
1,3,4	0.52
2,3,4	1.00
1,2,3,4	1.00

* Likely due to Braess's Paradox, i.e., Braess et al. "On a Paradox of Traffic Planning," *Transp Science*, 2005, 39(4): 446-450.

Glass Box Model: Illustrative Description

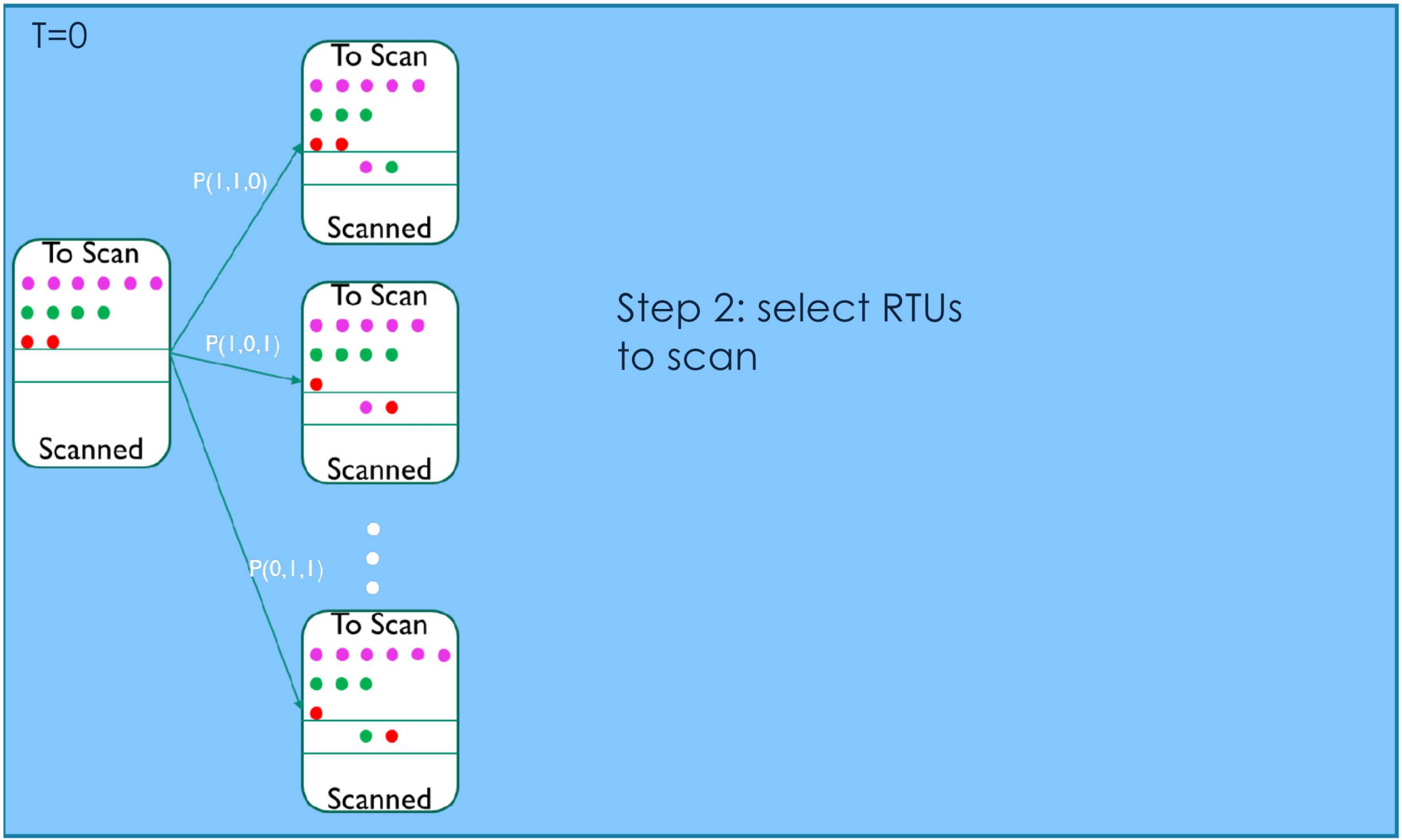
T=0



Step 1: set initial conditions

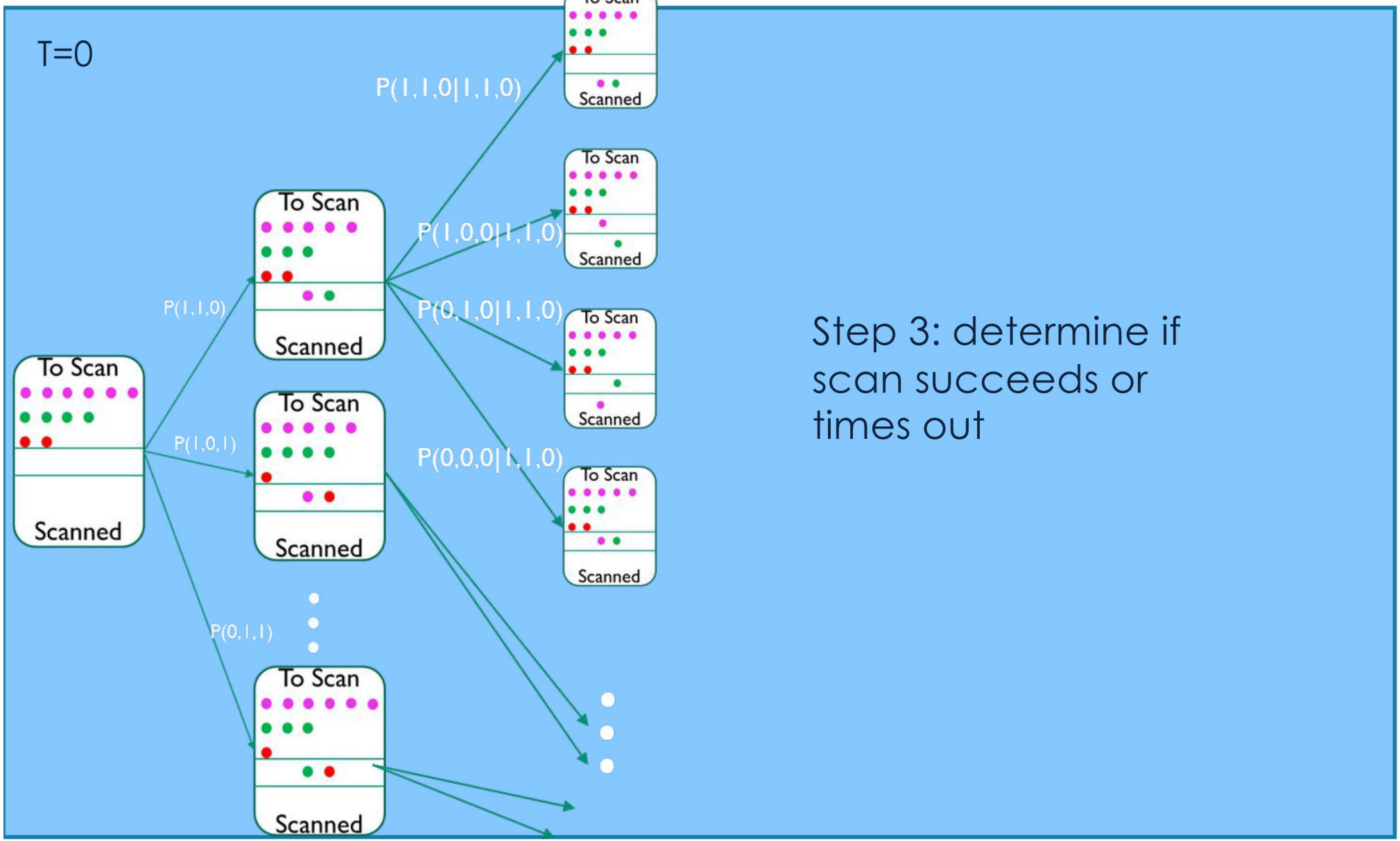
- = Filtered
- = Closed
- = Open

Glass Box Model: Illustrative Description



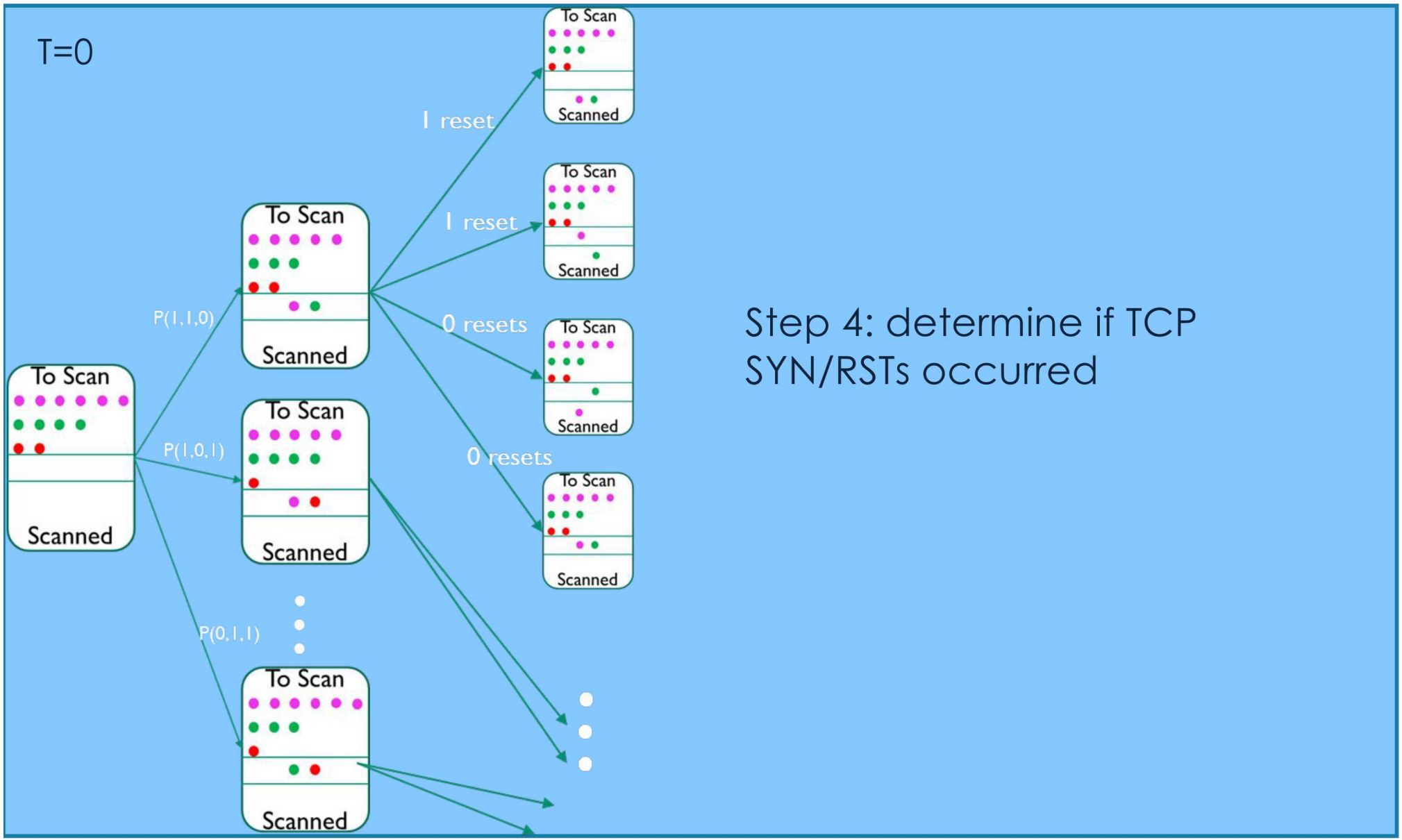
Step 2: select RTUs to scan

Glass Box Model: Illustrative Description



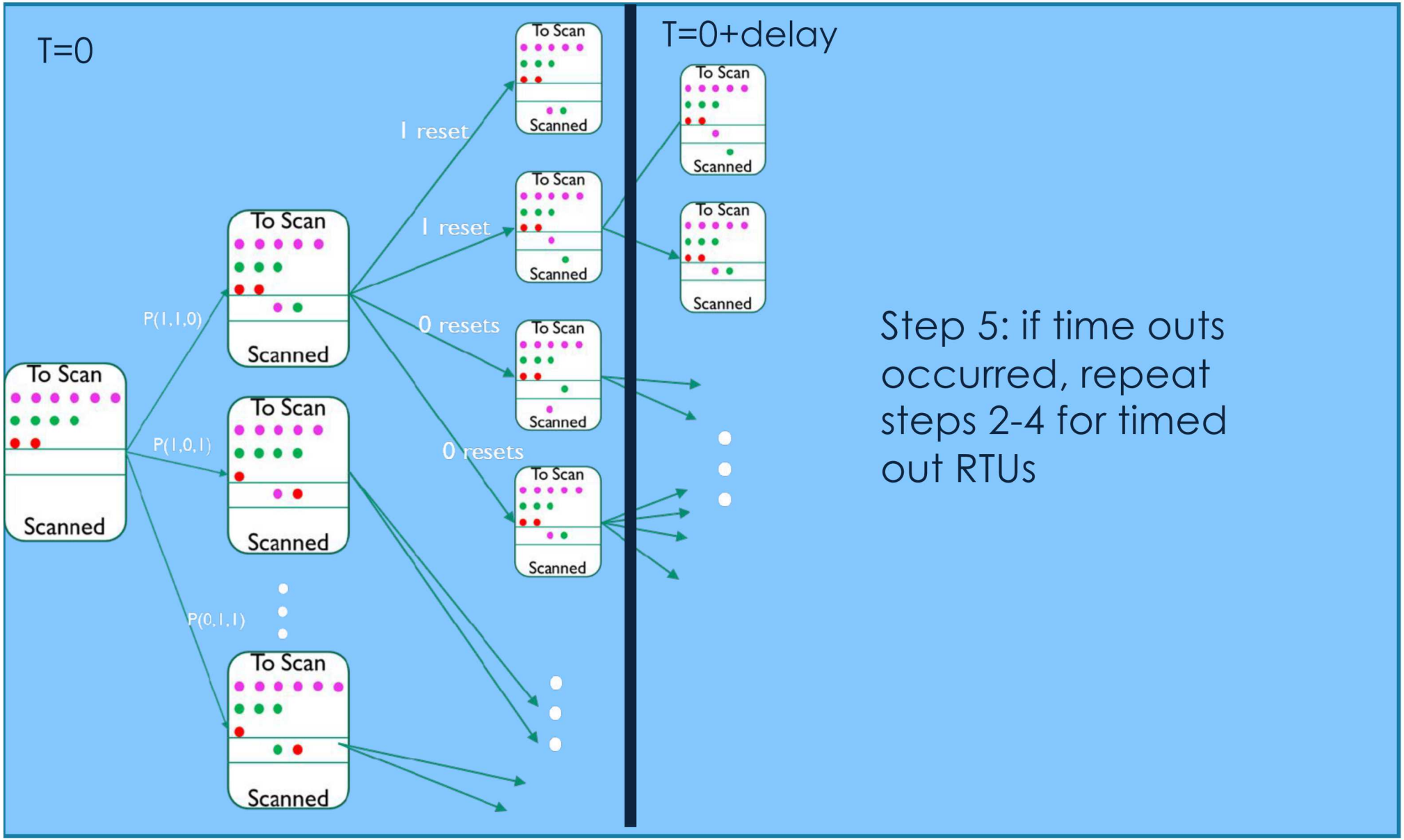
Step 3: determine if scan succeeds or times out

Glass Box Model: Illustrative Description



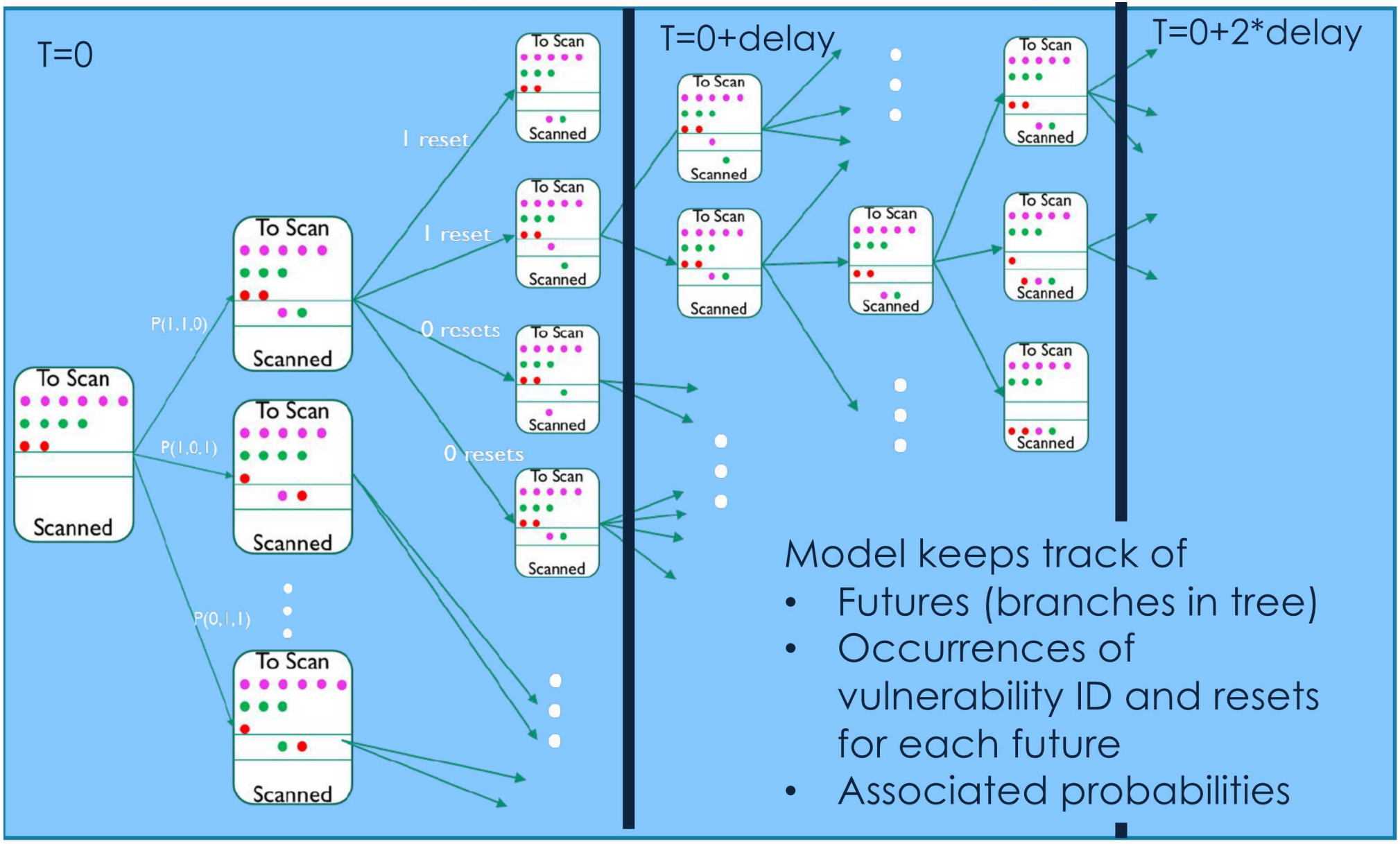
Step 4: determine if TCP SYN/RSTs occurred

Glass Box Model: Illustrative Description



Step 5: if time outs occurred, repeat steps 2-4 for timed out RTUs

Glass Box Model: Illustrative Description



Model keeps track of

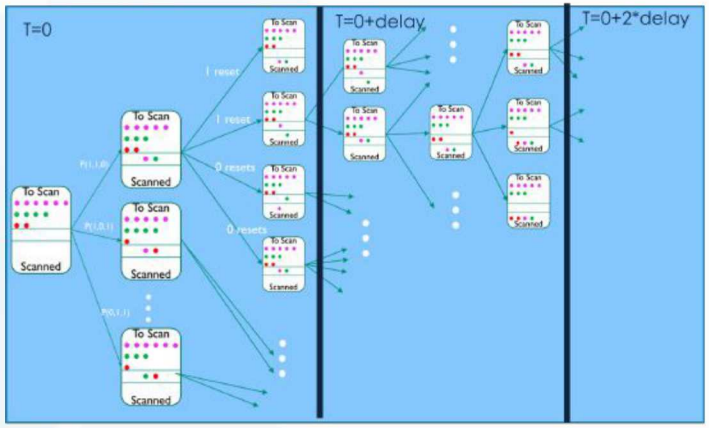
- Futures (branches in tree)
- Occurrences of vulnerability ID and resets for each future
- Associated probabilities

Glass Box Model: Implementation

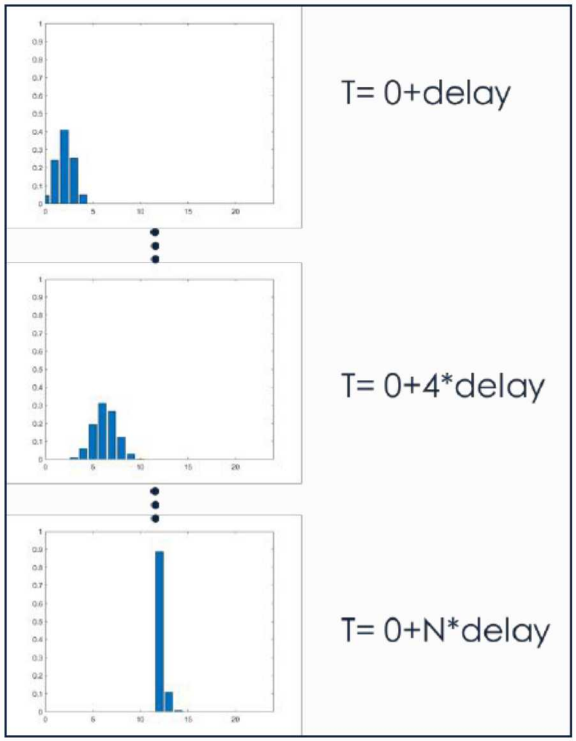
Inputs

Parameter Configuration

Analytical Solution (Not estimated via Monte Carlo Sims)



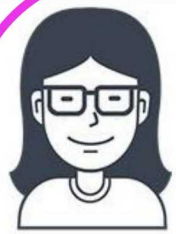
Probability Distributions at Discrete Times



Example Results

- System settings
  - 4 open (aka vulnerable) RTUs
  - 8 closed RTUs
  - 12 filtered RTUs
  - Probability of probe time out = 0.1
- NMap settings
  - Host group: 4
  - Scan delay: 10s
  - Max # of retries: 1
- Snort setting:
  - Low sensitivity
- Strategy:
  - *a priori*, attacker decides to wait for T seconds, and then attacks RTUs that have been identified by t=T.
  - If attacker detected before T seconds, attack fails and no load loss.

Questions



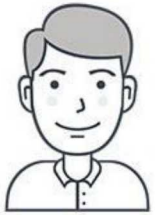
Adison the Engineer

- Analysis: for this scenario, can we estimate
  - Rate of vulnerable RTU identification?
  - Probability that the attacker is detected over time?
  - At which point during the scan should the attacker attack RTUs to maximize loss of load?

Wally the Attacker

- Validation: can we validate results from emulation experiments through comparison with glass box model estimates? And vice versa?

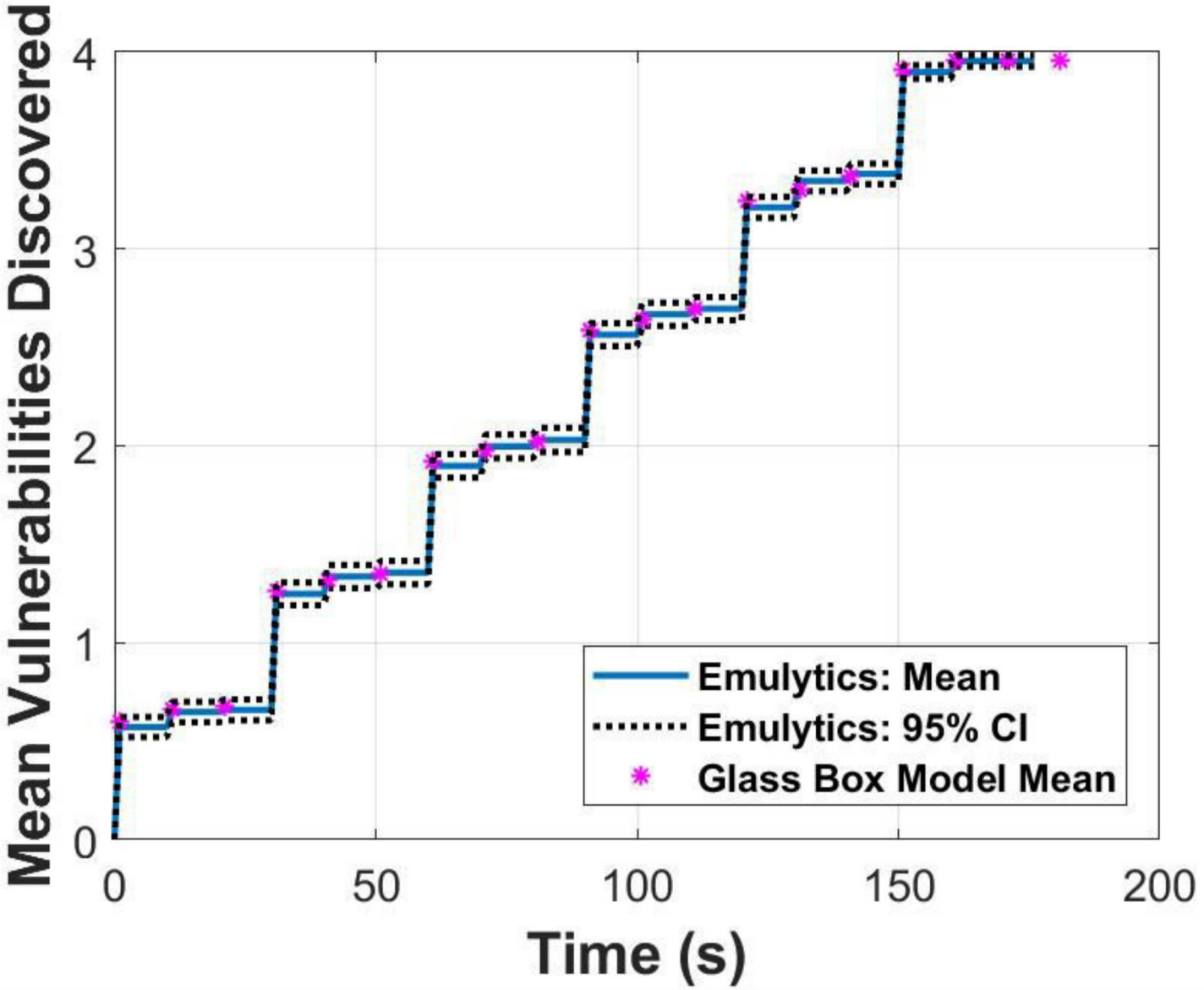
Dr. Turing the PI



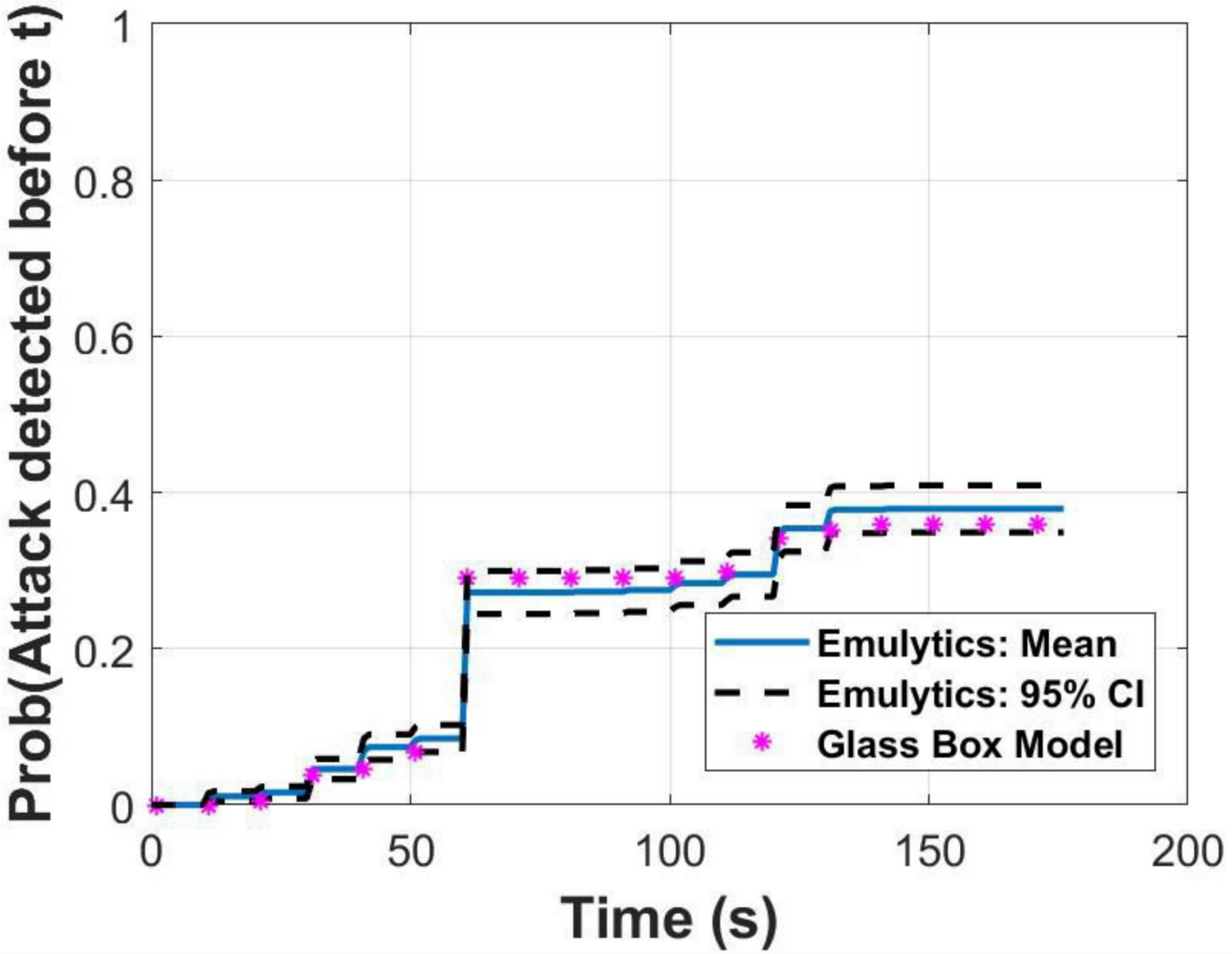
Leon the PM

- Practical consideration: how can we implement experiments in organized, efficient manner to capture potential uncertainties?

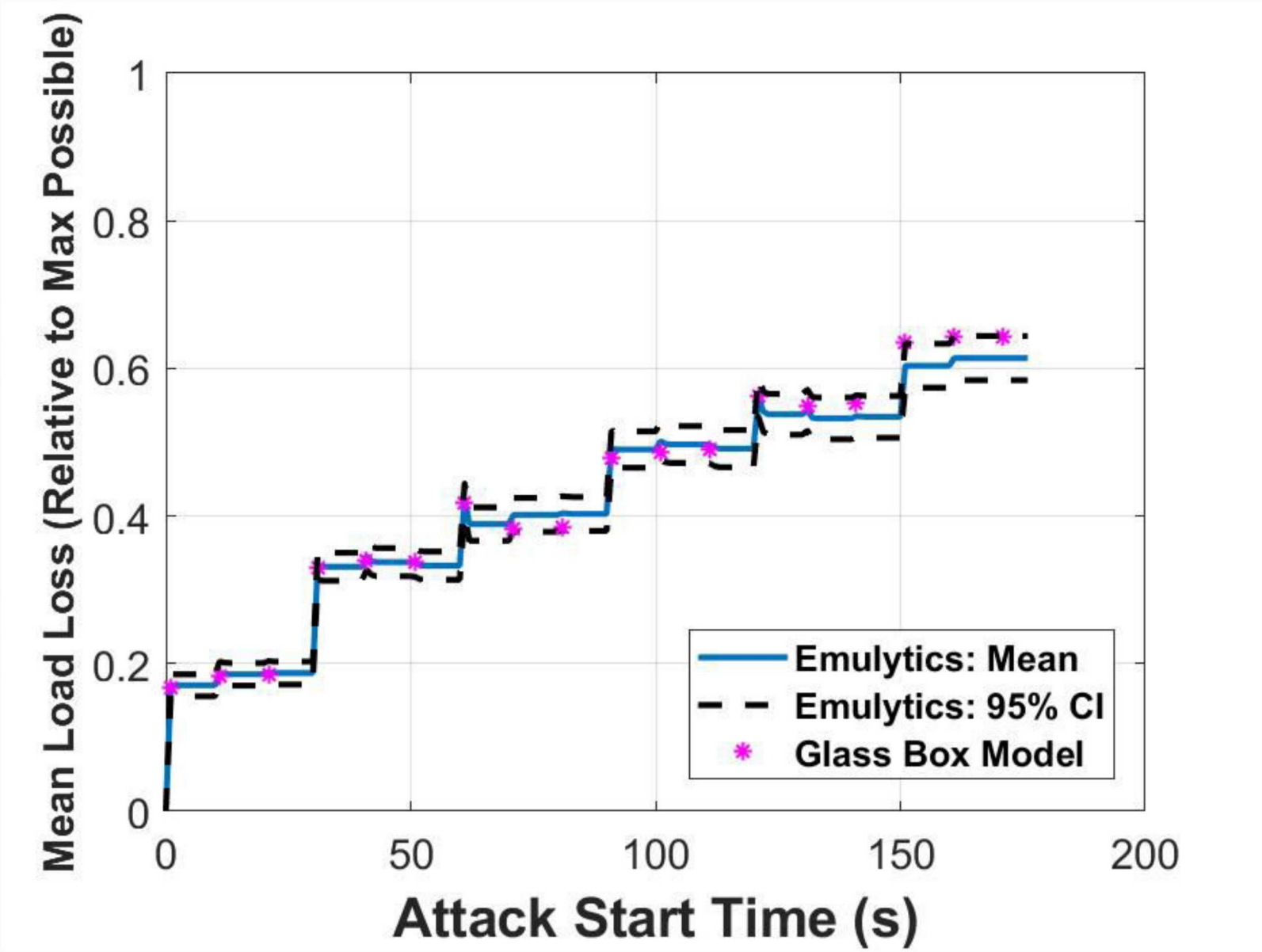
Results: Vulnerability Identification



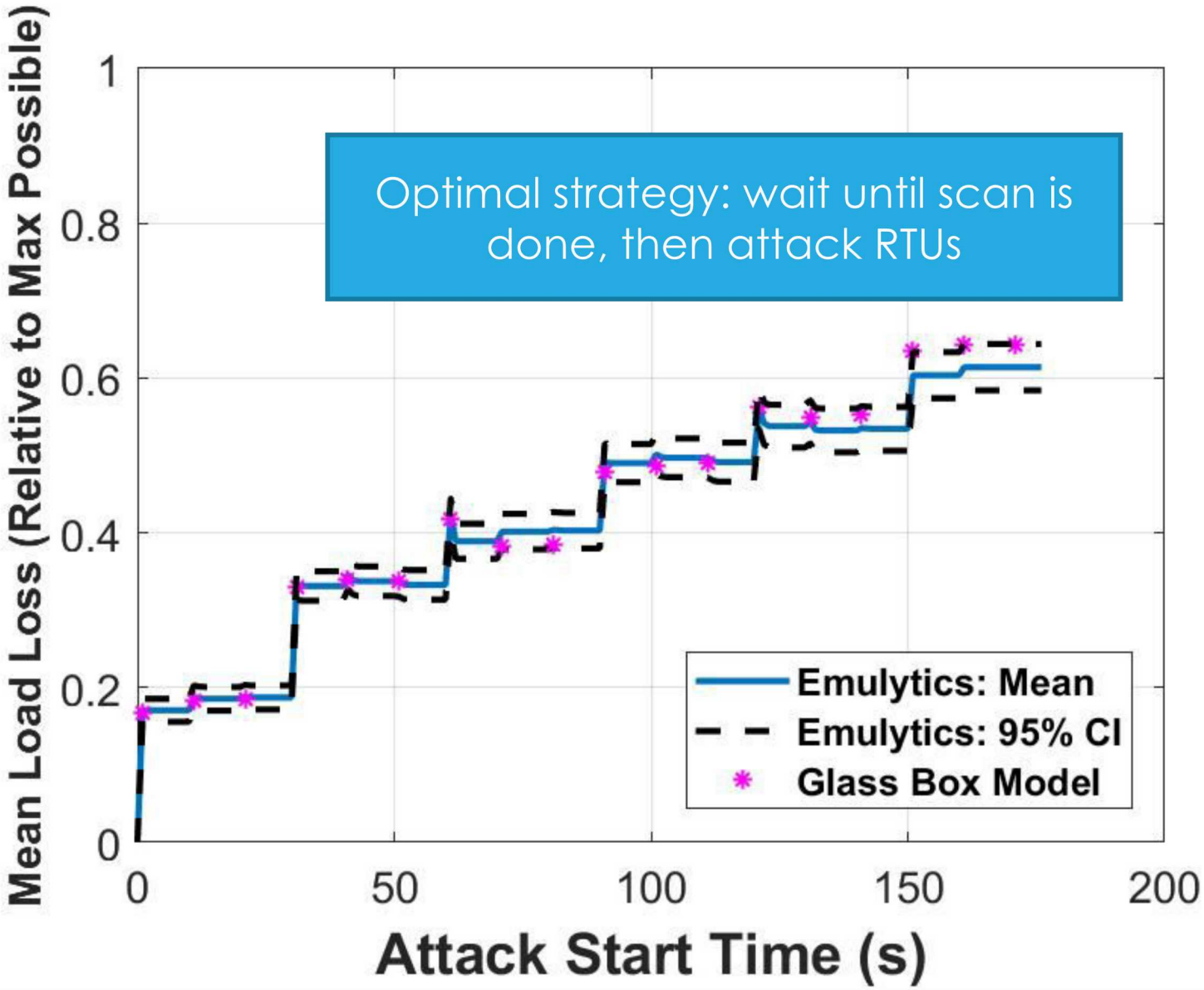
Results: Detection of Attacker



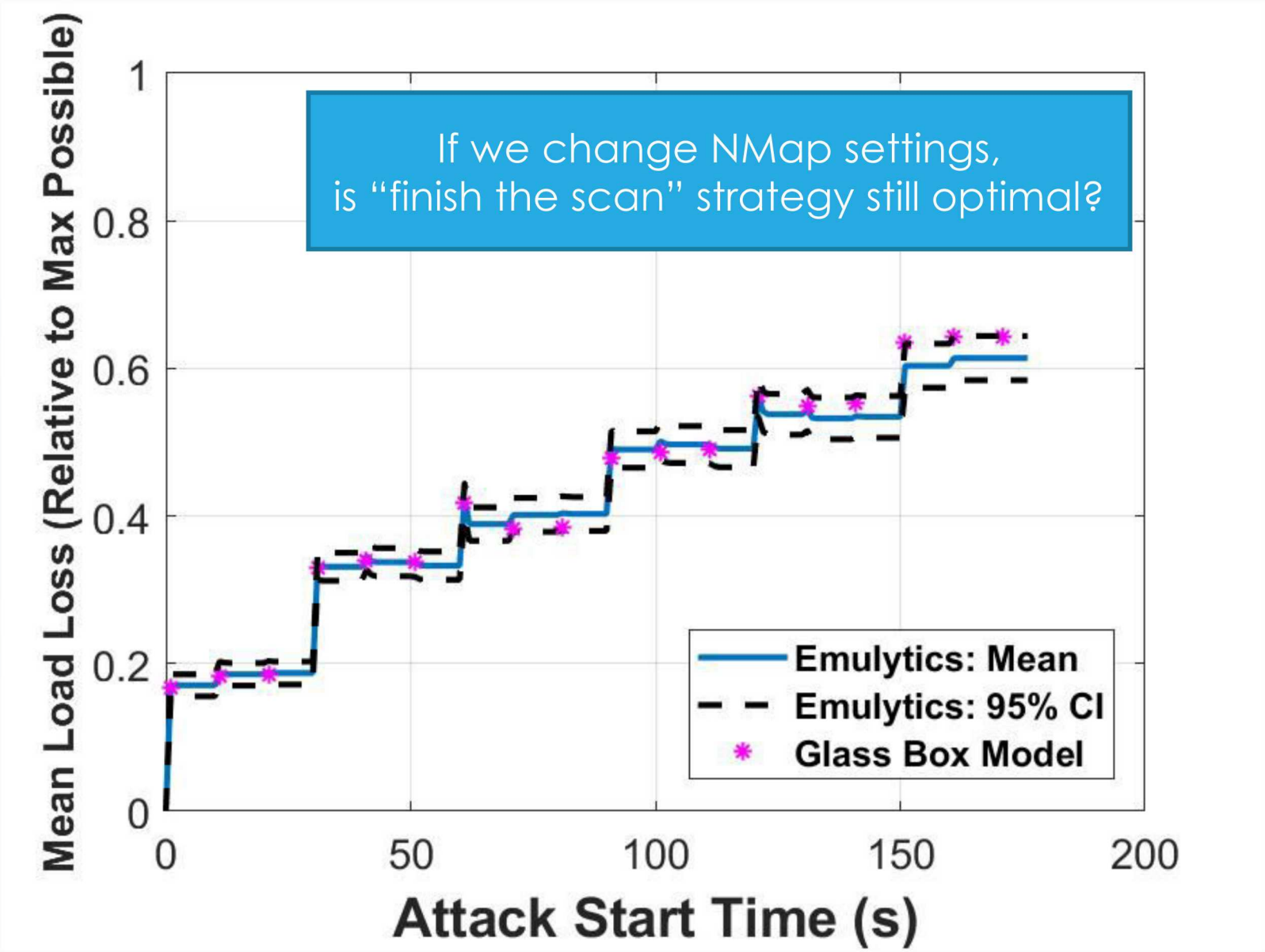
Results: Load Loss



Results: Load Loss



Results: Load Loss



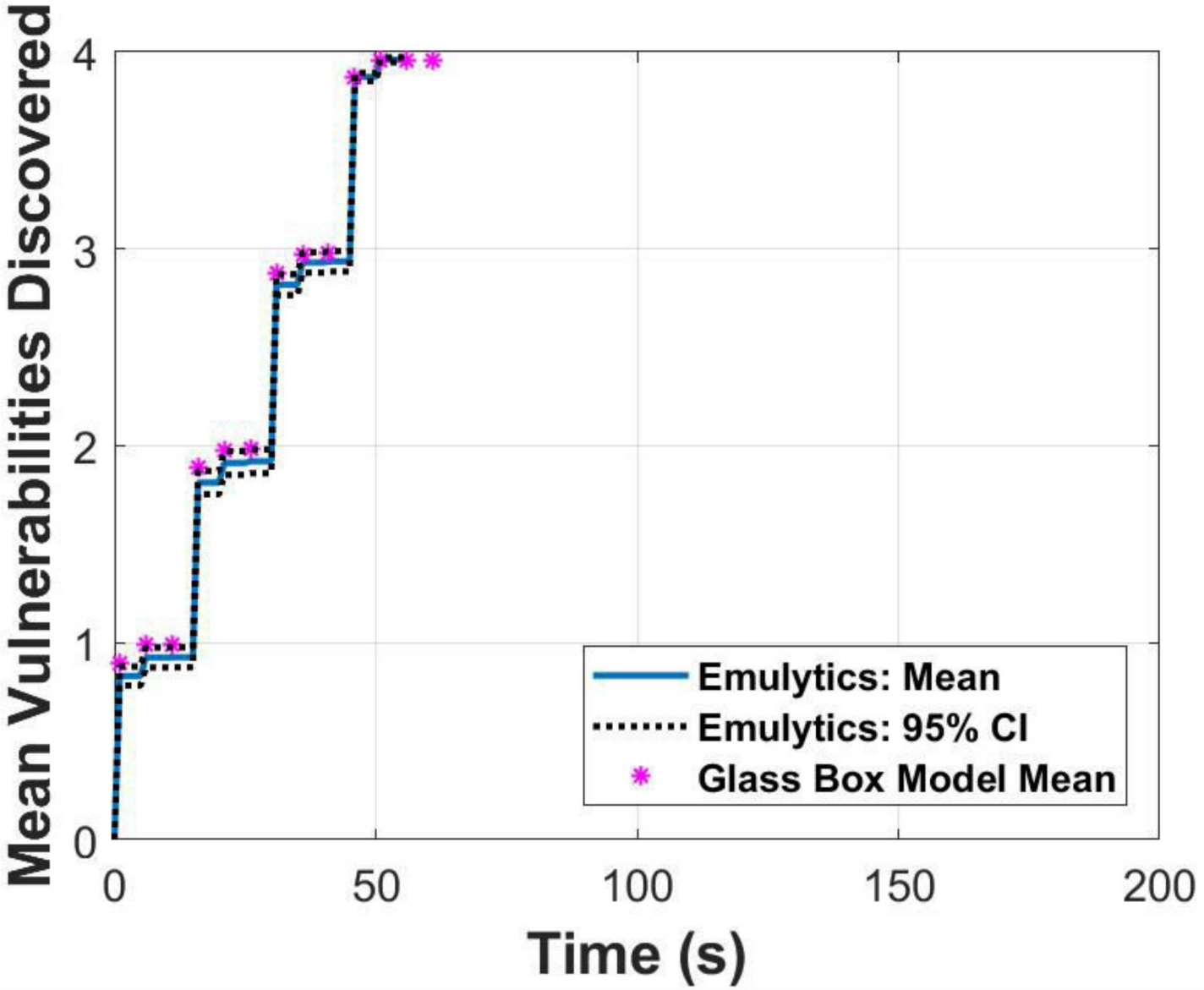
Example Results

- System setting
  - Probability of probe time out = 0.1
- NMap settings
  - Host group: 6
  - Scan delay: 5s

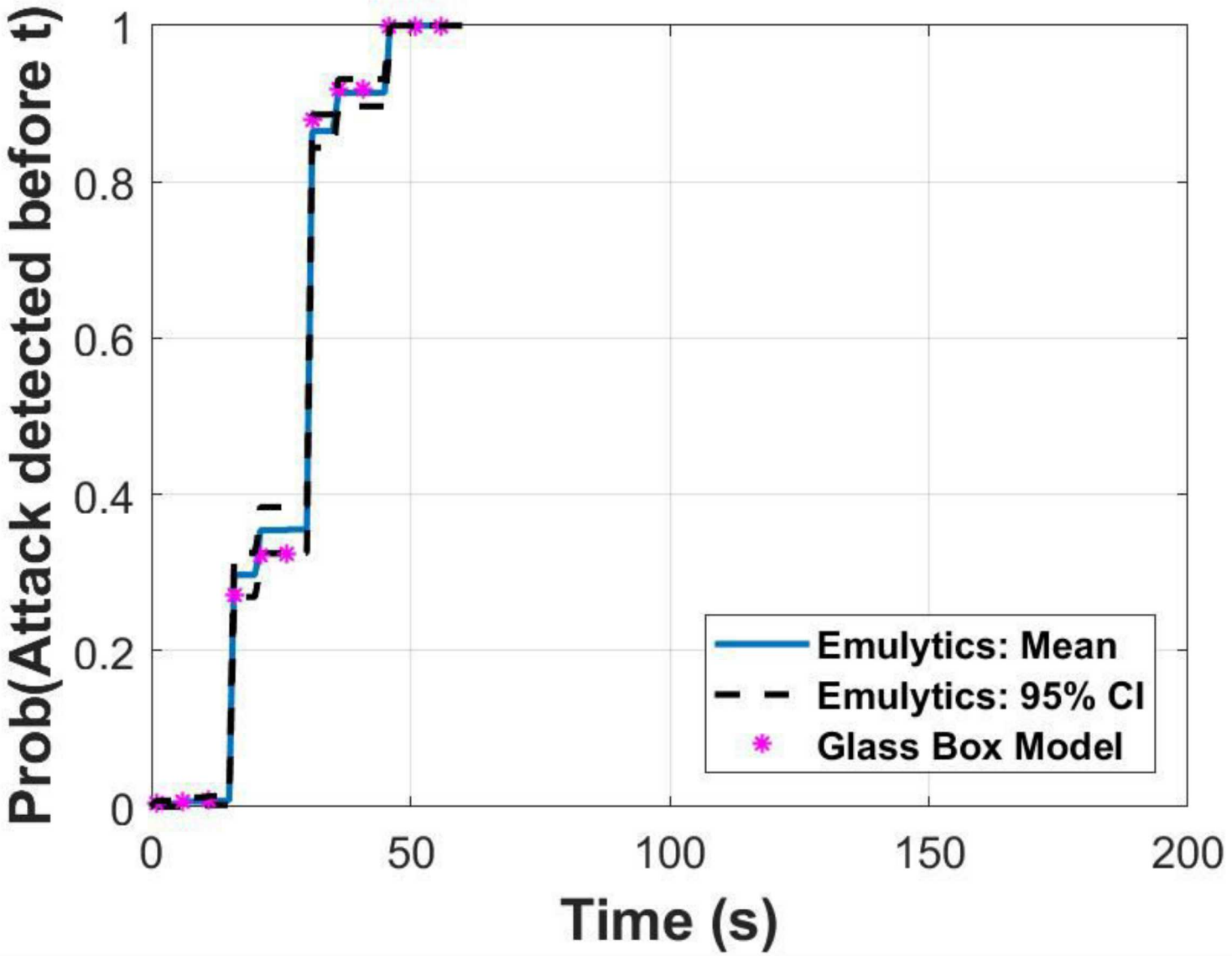
Faster scan, higher chance of detection

- Snort setting:
  - Low sensitivity
- Strategy:
  - *a priori*, attacker decides to wait for T seconds, and then attacks RTUs that have been identified by t=T.

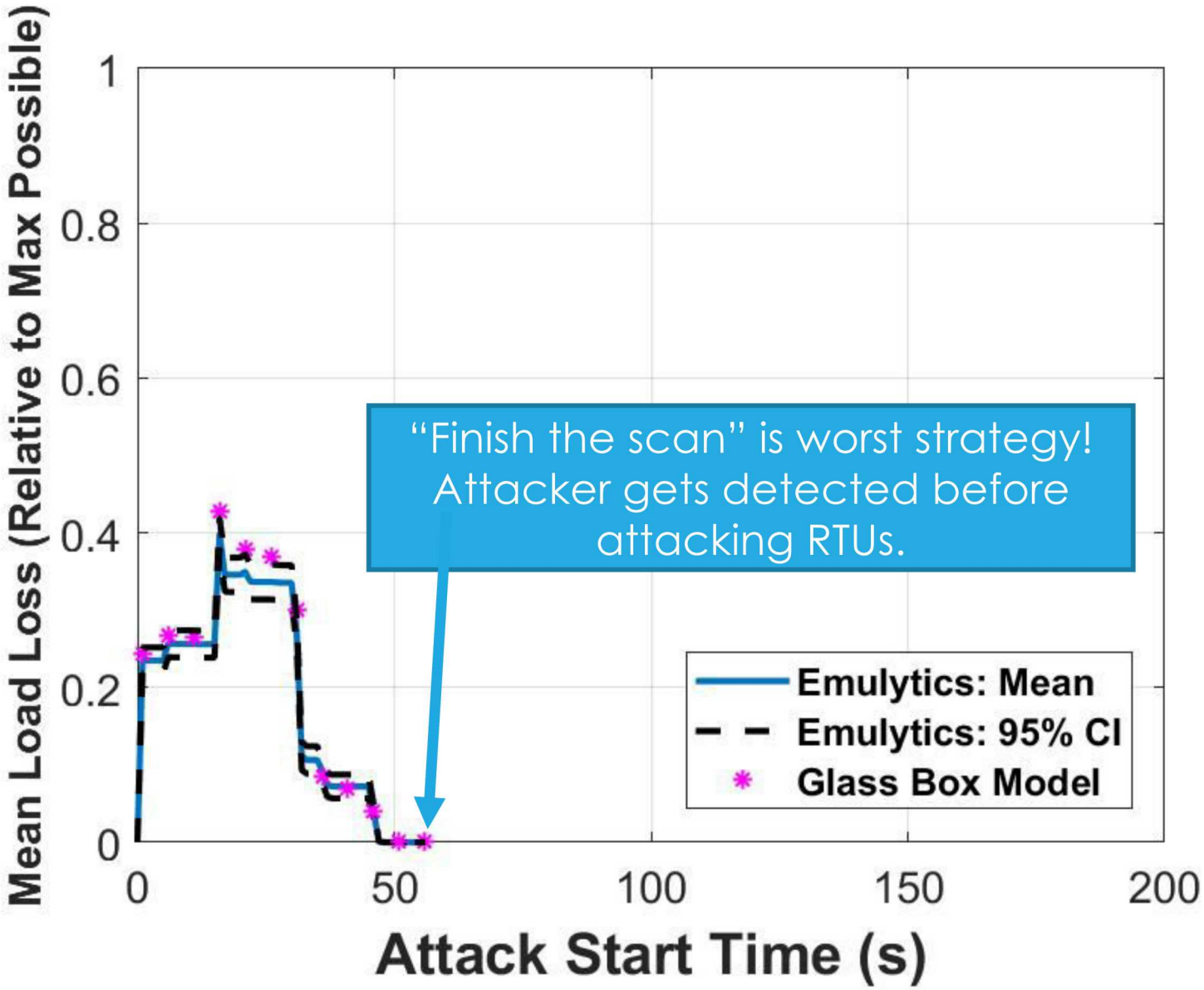
Results: Vulnerability Identification



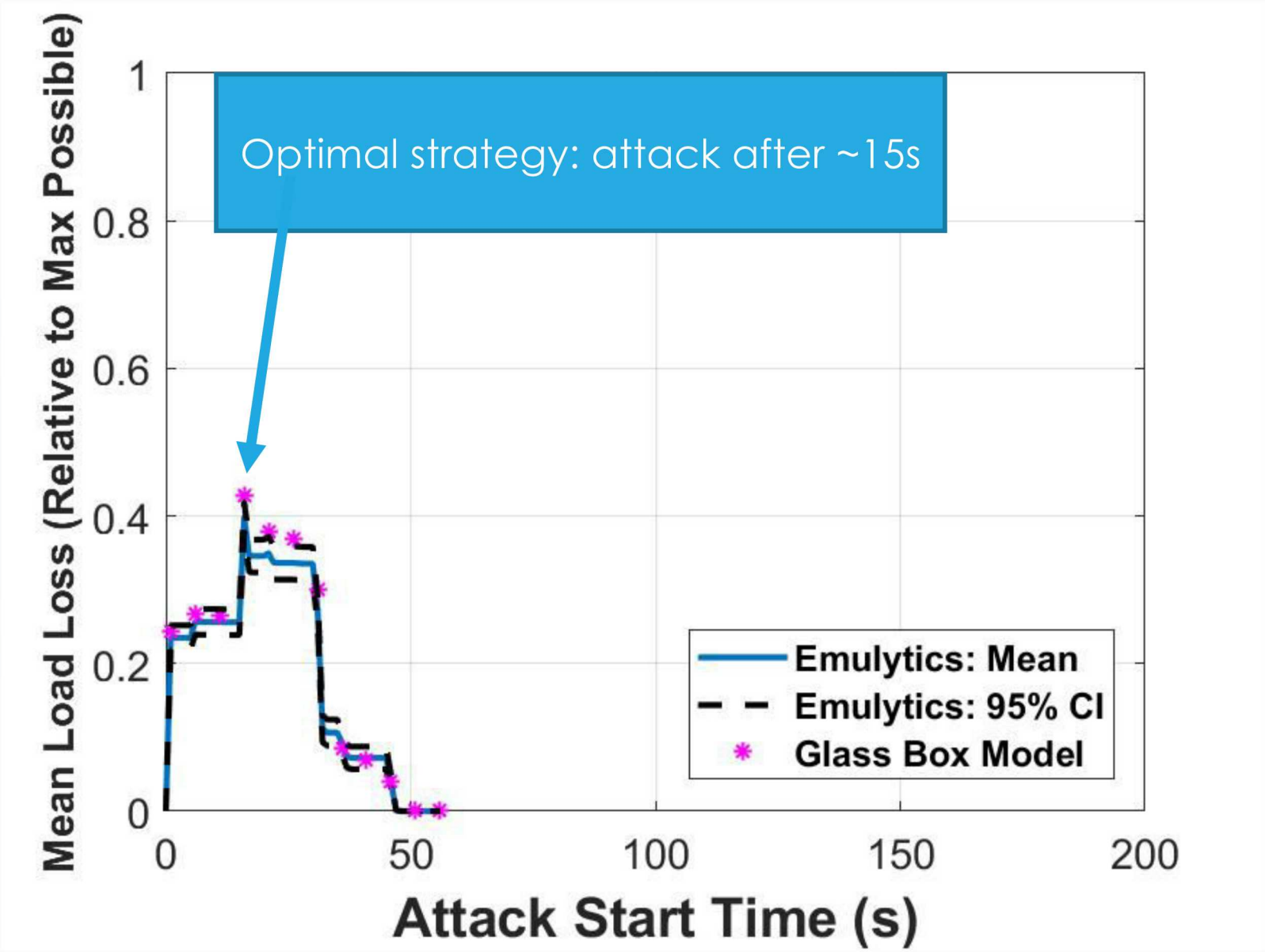
Results: Detection of Attacker



Results: Load Loss



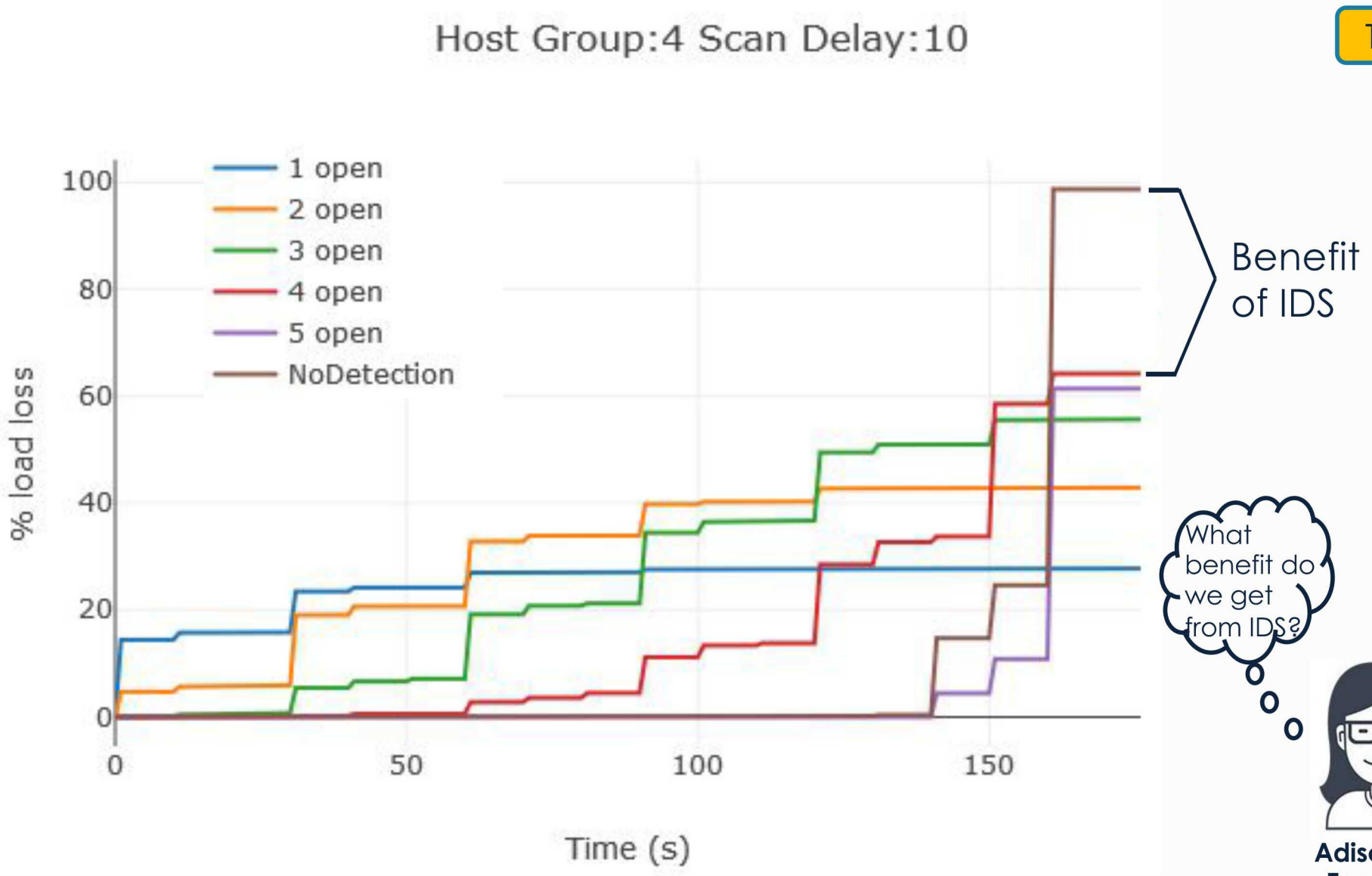
Results: Load Loss



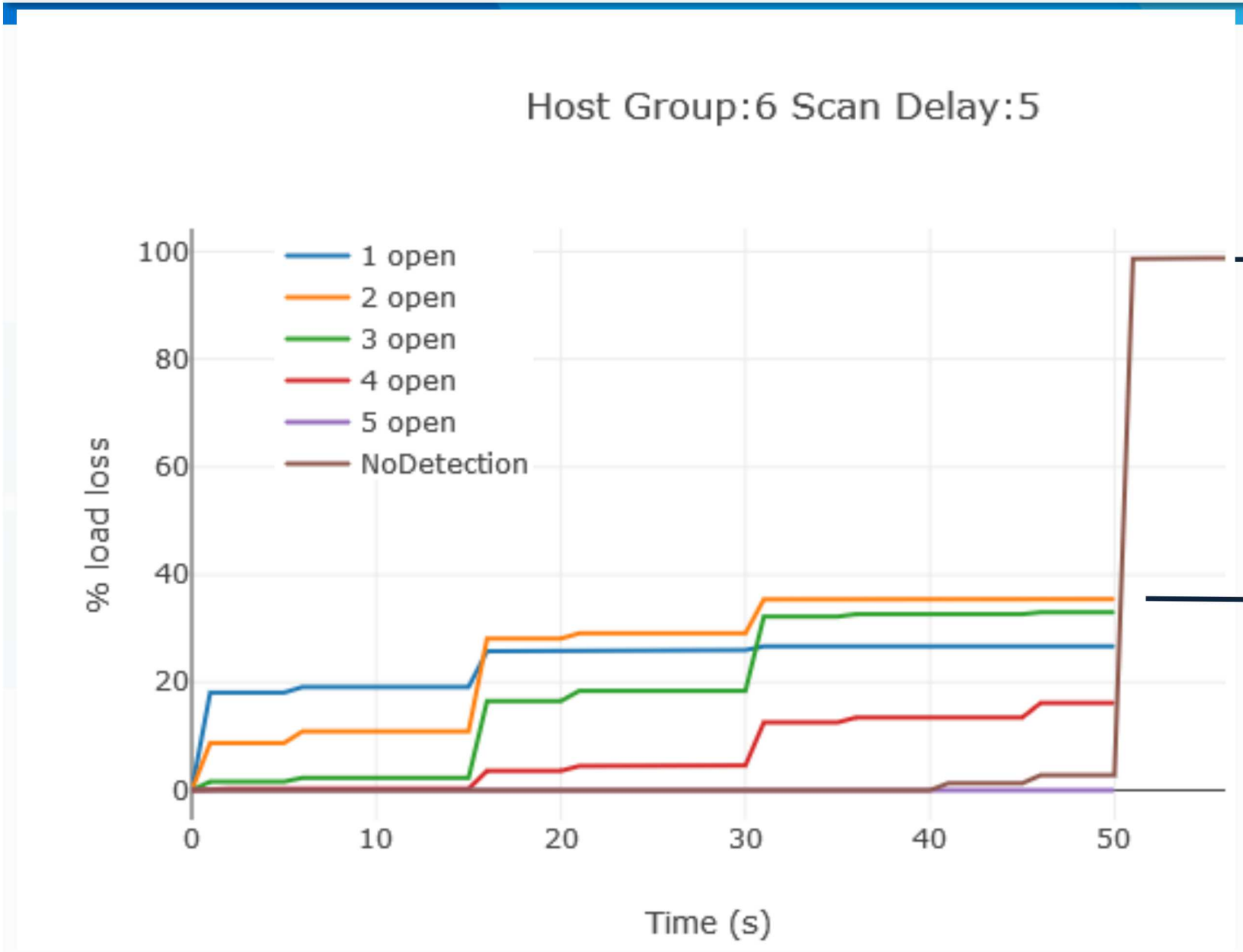
Example Results

- System setting
  - Probability of probe time out = 0.1
- NMap settings
  - Host group: 4
  - Scan delay: 10s
- Snort setting:
  - Low sensitivity
- Strategy: attacker uses feedback from scans to determine when to attack
  - Attack RTUs after finding N vulnerabilities

Results: Load Loss vs. Strategy



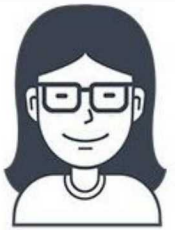
Results: Load Loss vs. Strategy



TC

Benefit of IDS

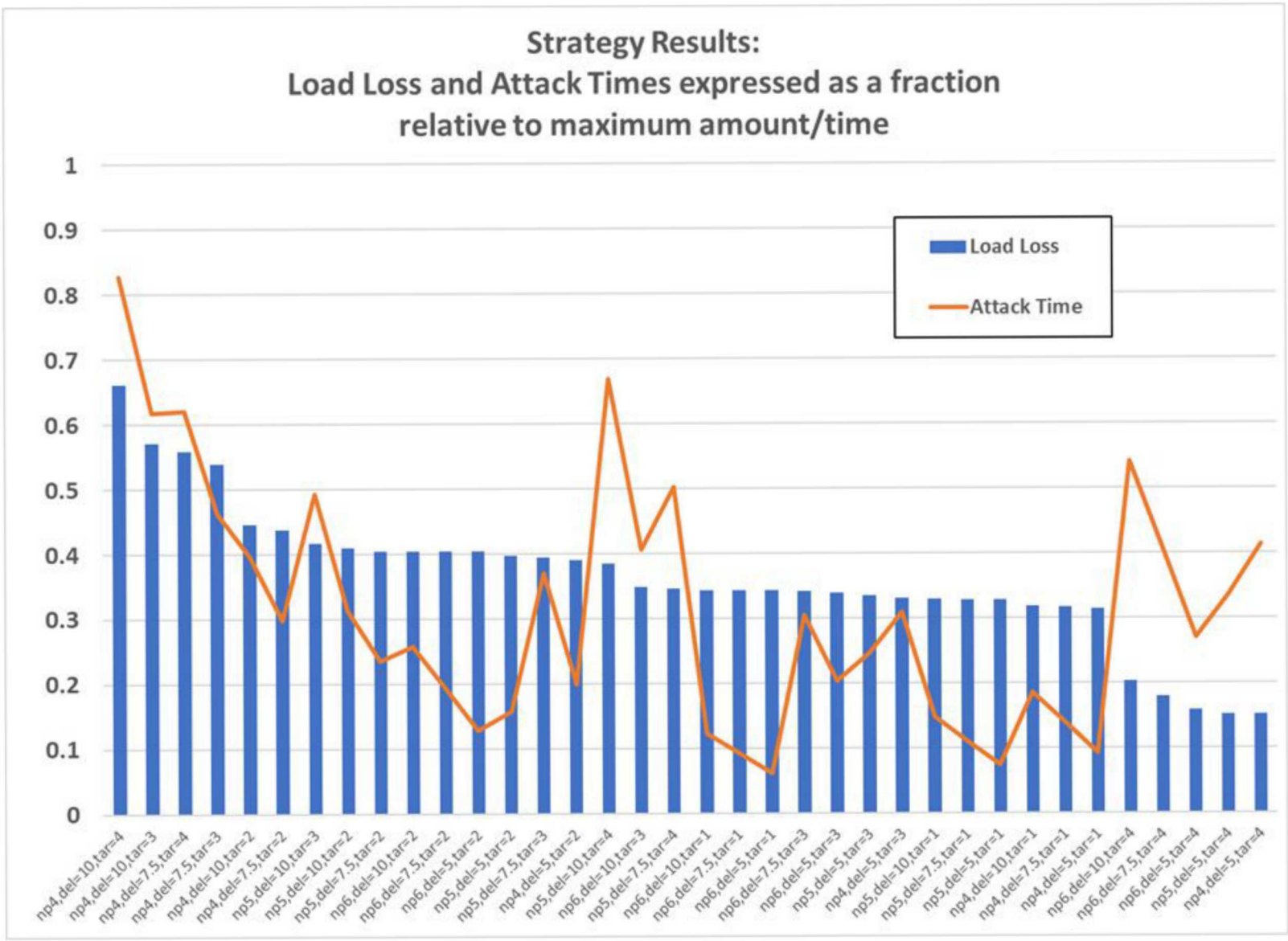
What benefit do we get from IDS?



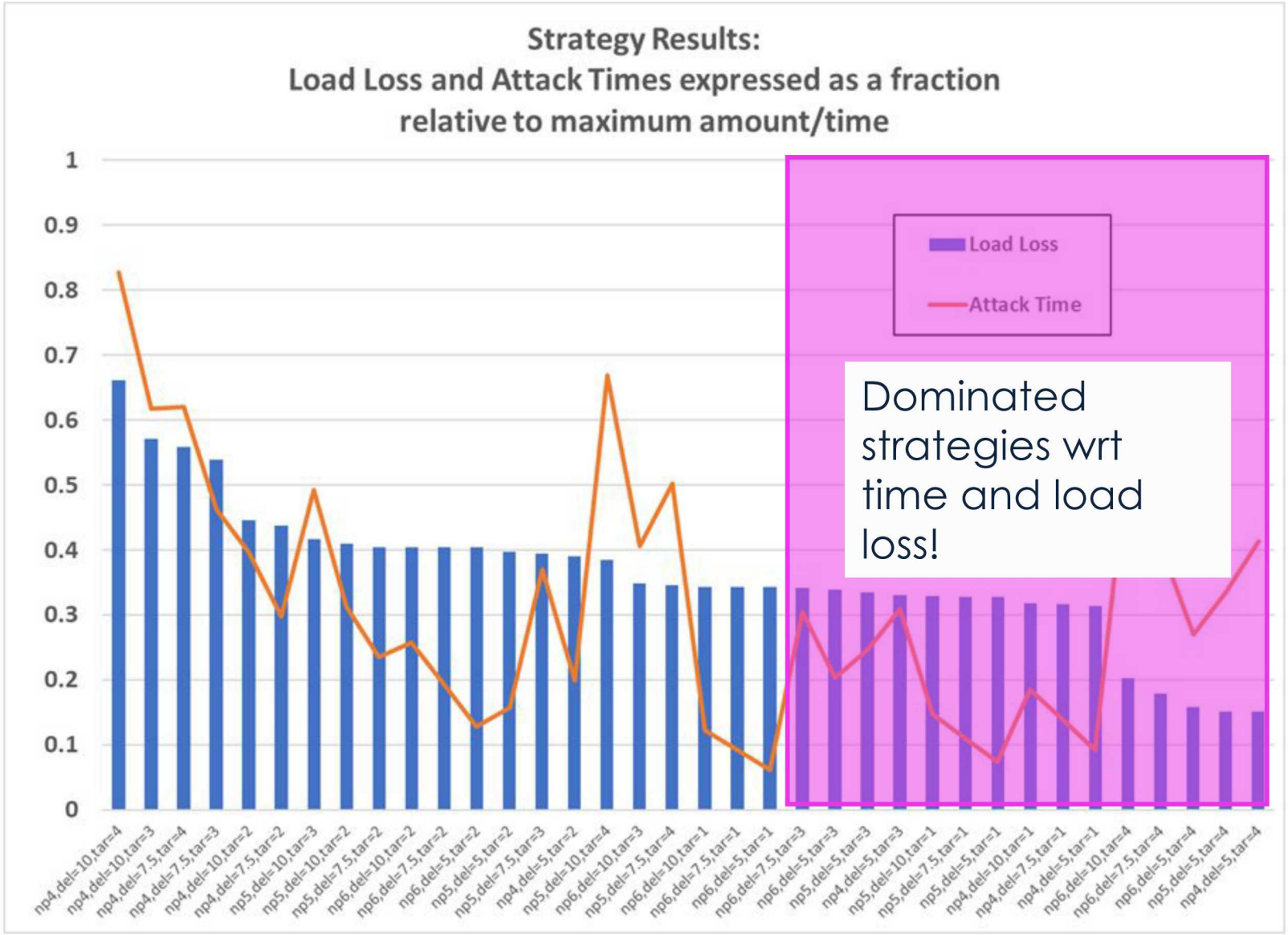
Adison the Engineer

Results: Load Loss vs. Strategy

TC

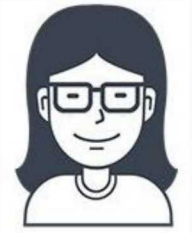


Results: Load Loss vs. Strategy



TC

Questions



Adison the Engineer

- Analysis: for this scenario, can we estimate
  - Rate of vulnerable RTU identification?
  - Probability that the attacker is detected over time?
  - At which point during the scan should the attacker attack RTUs to maximize loss of load?

Wally the Attacker

- Validation: can we validate results from emulation experiments through comparison with glass box model estimates? And vice versa?

Dr. Turing the PI

Leon the PM

- Practical consideration: how can we implement experiments in organized, efficient manner to capture potential uncertainties?

Experimental Set Up: SCORCH

- SCORCH: **SC**enario **ORCH**estration tool for minimega

ST

```
root@en189:~# scorch -h
usage: scorch [-h] [--run_name RUN_NAME] [--namespace NAMESPACE] configuration

Securetk.emulytics scenario orchestration tool

positional arguments:
  configuration         Name of scenario configuration to run

optional arguments:
  -h, --help            show this help message and exit
  --run_name RUN_NAME   Name of scenario run
  --namespace NAMESPACE
                        Name of namespace to run against
```

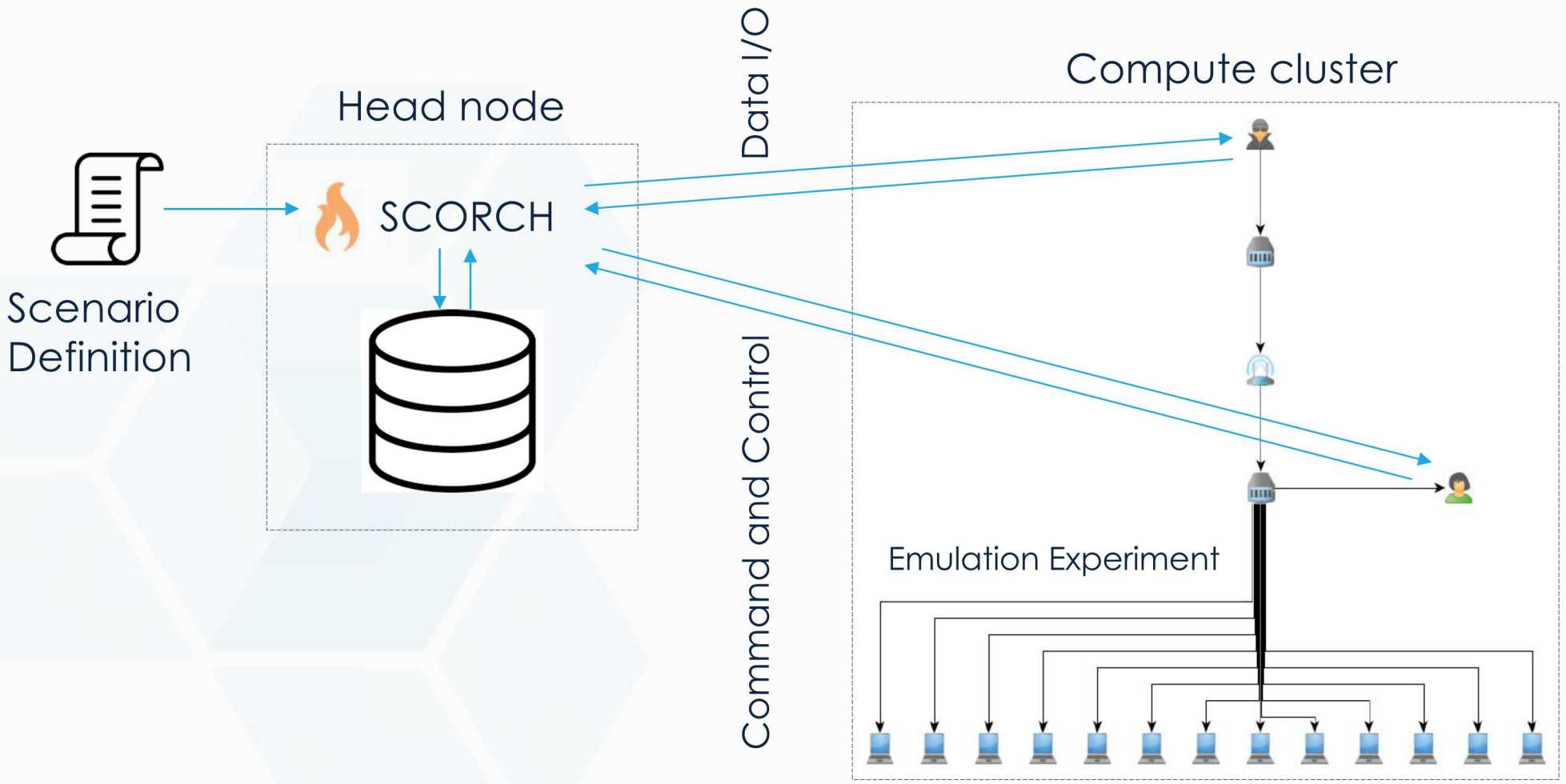
- Implements a simple "scenarios" scripting language
- Facilitates experimental data I/O to and from the emulated network environment
  - Inspired by Distributed Experiment Workflows (DEW)[1]
- Enables rapid development of repeatable scenarios
- Modular scenario "components" promote reuse

[1] Jelena Mirkovic, Genevieve Bartlett, and Jim Blythe. 2018. DEW: distributed experiment workflows. In *Proceedings of the 11th USENIX Conference on Cyber Security Experimentation and Test* (CSET'18). USENIX Association, Berkeley, CA, USA, 4-4.

Experimental Set Up: SCORCH

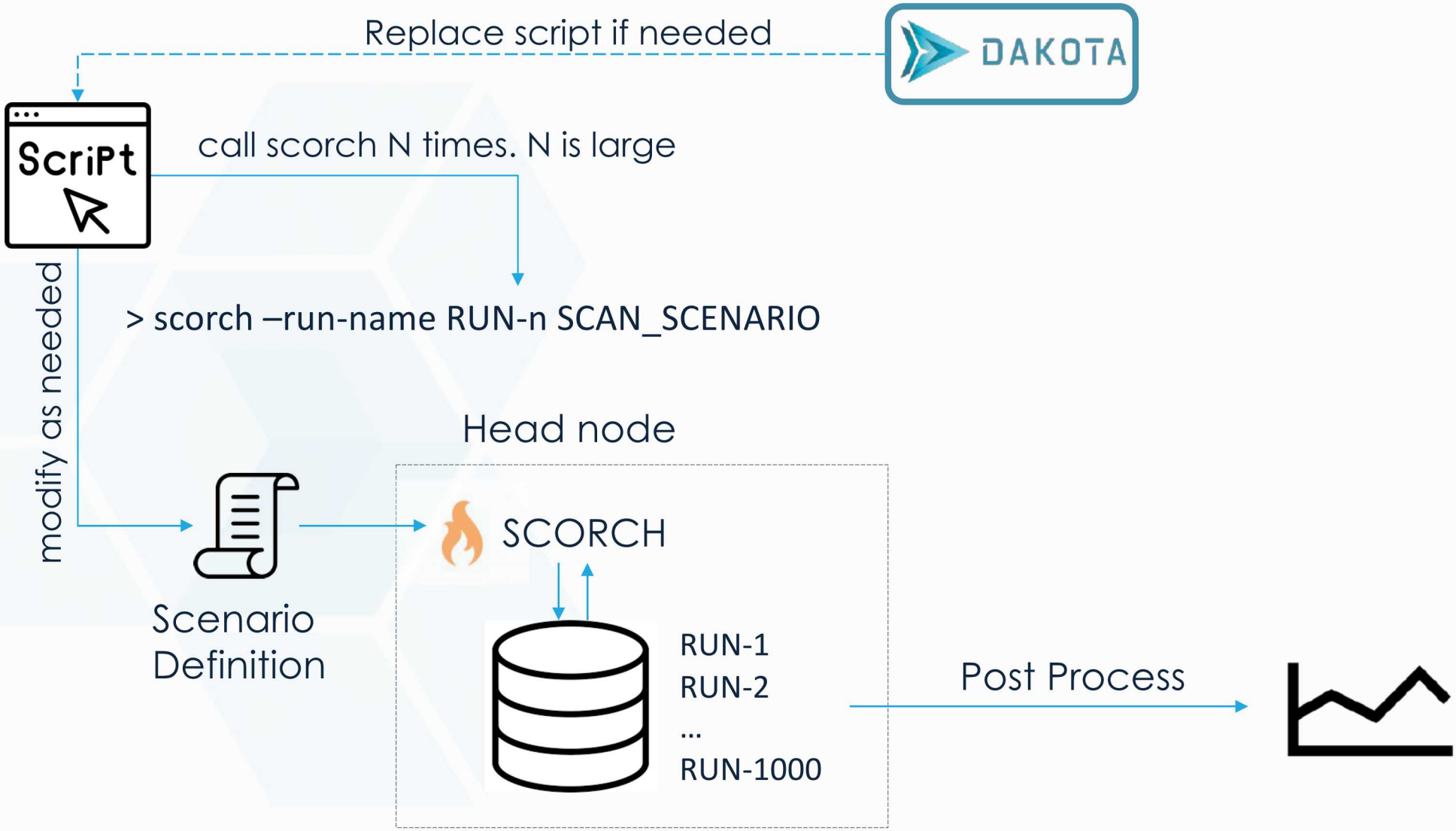
- Projects that use/considering using SCORCH:
  - Advanced C2 threat modeling
  - Behavioral analytics for ICS
  - Resilience analysis for energy systems
- Project needs:
  - Running large numbers of repeatable scenarios in emulated environments
  - Data collection from emulation experiments
    - Validation
    - Training data
    - Exploratory analysis

Experimental Set Up: SCORCH



> scorch –run-name RUN-1 SCAN_SCENARIO

Experimental Set Up: SCORCH



Insights and Progress

- Improved Emulytics infrastructure for repeating large # of experiments
  - Efficient data extraction and analysis

	Initially (manual/serial)	Currently (parallel)
Perform Emulytics Runs	~10 days	3 hrs
Process Data	days	hrs - mins
Restarts	days	far fewer

  - Improved consistency for repetition of Emulytics experiment
  - Modular design ought to work well with other studies
- Threat model considered both technology and strategic elements
  - COTS technologies: NMap, Snort
  - Strategy tradeoffs: benefits/drawbacks for attack strategies
- Study catalyzed collaboration with UQ team

Insights and Progress

- Benefits from co-development of Emulytics and glass box models
  - Emulytics experiments helped identify undocumented NMap behaviors to include in the glass box model
  - Glass box model was used to explore space of parameters before performing Emulytics runs
  - Cross-validation and verification
- Glass box model provides efficiencies under some conditions

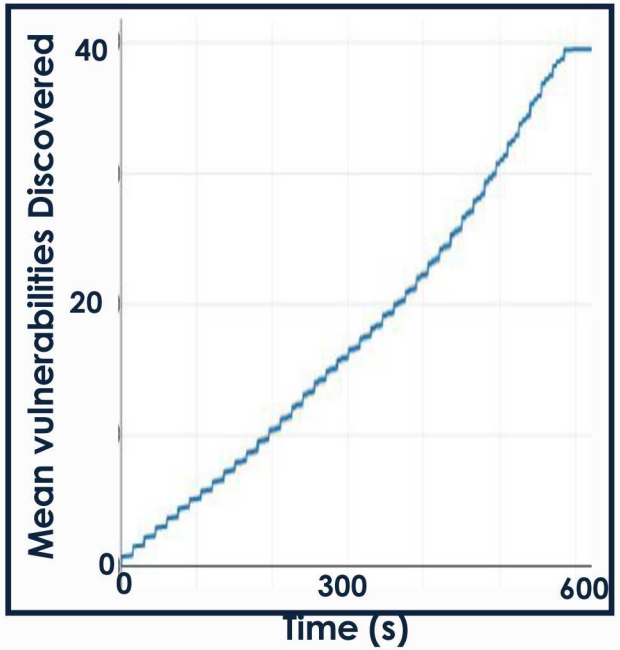
4 open, 8 closed, 12 filtered	np=4	np=5	np=6
Scanning (normalized)	~4s	<1 s	<1 s
detection: 1 delay/threshold setting	~12 hrs	~4.5 hrs	~0.5 hrs
detection: additional delay/threshold	~ 6 hrs	~0.8 hrs	~0.1 hrs

- Glass box model differs from scanning-/security-related efforts e.g.
  - Toutonji et al. 2012; Chen & Ji 2005 (many others): using epidemic propagation models to model spread of internet worms and malware
  - Turner & Joseph 2017; Huang et al. 2012: analyzing/improving Snort
  - Alpcan & Basar 2005: game theory/optimal control analysis of IDS
  - Wang et al. 2012 : attack-defense stochastic game net

Unexpected Results and Next Steps

- Expected "simple" example showed surprising complexity
- Challenges comparing "discrete time" model with "continuous time results"
- Next step: draft manuscript for publication
- Possible next steps
  - Scaling up
  - Add "background traffic"
  - Test/validate on non-emulated network
  - Explore strategy evaluations for this problem
    - Have some ideas using partially observable Markov decision process models
    - Good opportunity for further collaboration with optimization team
  - Consider modeling and analysis for other portions of the kill-chain, e.g. command and control

Scaling Up: 10x Expt.



Total RTUs =240
Open = 40
Closed = 80
Filtered =120

Feedback

- Feedback on all aspects of the effort is appreciated
- Feedback on following topics would be especially appreciated
  - Suggested publication forums
  - Ideas on next steps
  - Further developments to SCORCH

Questions?

SAND2019-XXXX

LDRD

Laboratory Directed Research and Development

SECURE Uncertainty Quantification Thrust

Team Members:
Laura Swiler
Bert Debusschere
Gianluca Geraci
Jonathan Crussell
Erin Acquesta

Presenter:
Laura Swiler

U.S. DEPARTMENT OF ENERGY

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.

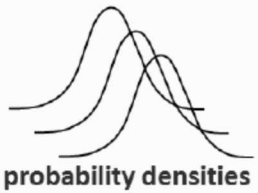
SECURE: **S**cience and **E**ngineering of **C**ybersecurity by **U**ncertainty quantification and **R**igorous **E**xperimentation

The Goal: Bring rigor into cyber experimentation

UQ Team: Develop and deliver approaches which allow uncertainty quantification to be performed on Emulytics efficiently.

Forward UQ: propagate uncertainties on inputs to uncertainty on predictions

Uncertainty in input variables *u*



Emulytics Model *f(u)*

Statistics on output *s(f(u))*

What does success look like?

STEPS **Year 1**

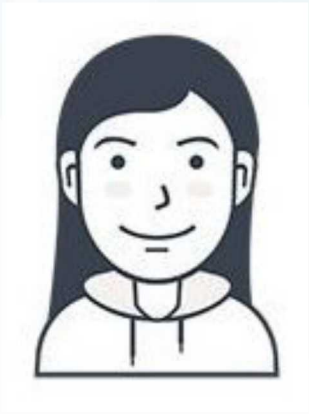
1. Demonstrate that we can sample Emulytics models reproducibly across platforms
   - Establish interface to Emulytics models for running ensembles
   - Sampling strategies
   - Characterization of input distributions
2. Validate a specific Emulytics problem (e.g. a particular network and threat)
3. Develop methods that can perform the forward UQ problem more efficiently
   - Sampling of discrete variables, experimental design
   - Dimension reduction
   - Multi-fidelity approaches
4. Demonstrate a full UQ workflow that is generalized over multiple threats and networks at scale.

Outline

- Analysis of the Exemplar
  - Sensitivity with respect to host group size/delay
  - Uncertainty with respect to the attacker strategy
  - Reproducibility of emulation runs across platforms and in parallel runs
- Research Thrusts
  - Multifidelity Uncertainty Quantification
  - Discrete uncertainties
  - Dimension Reduction
- FY2020 Plans
- Deep dive into Multifidelity

Uncertainty Analysis of Exemplar

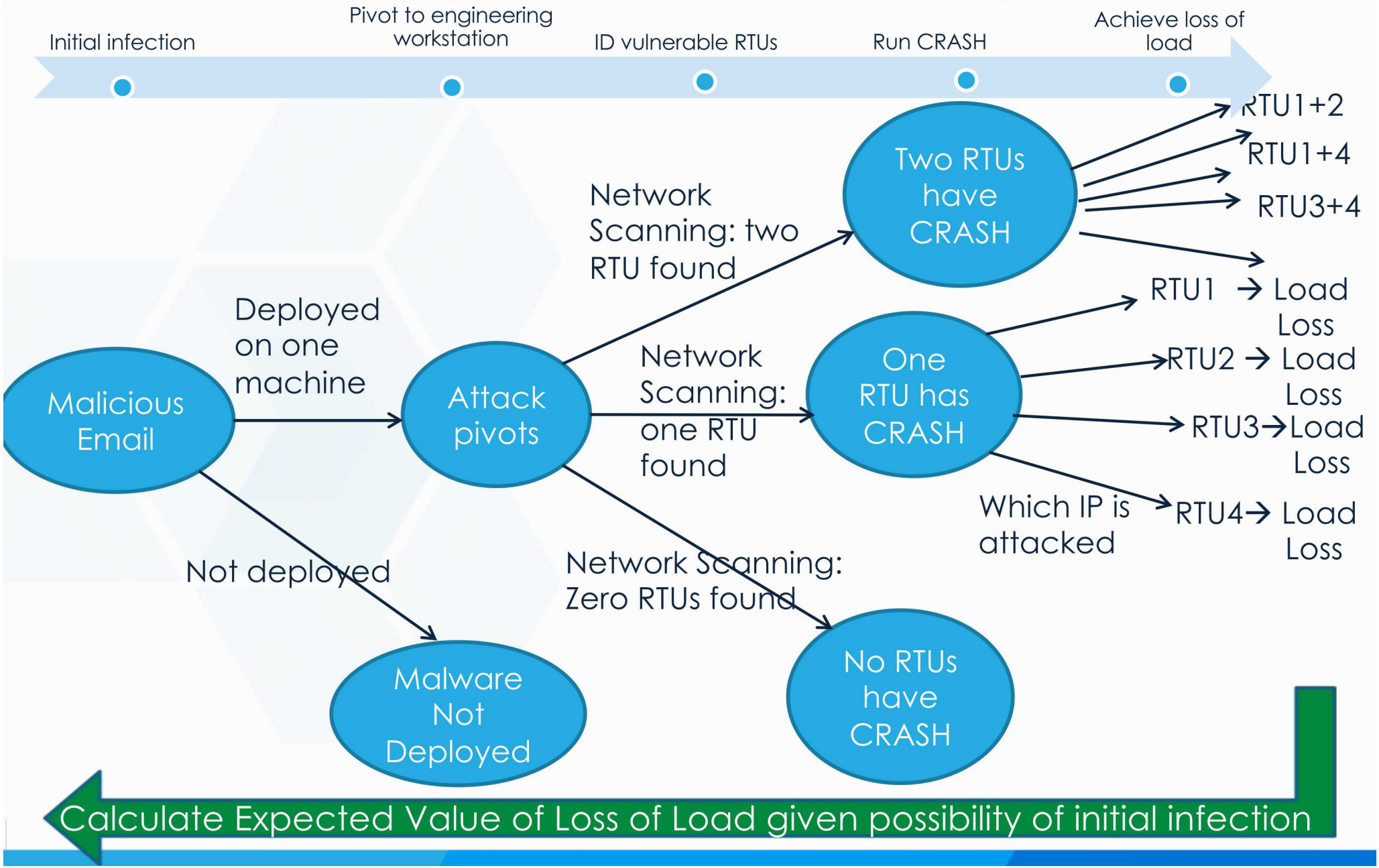
Alice, Designer



Attacker

Captain Howard, Defender

Analysis of Exemplar Uncertainties

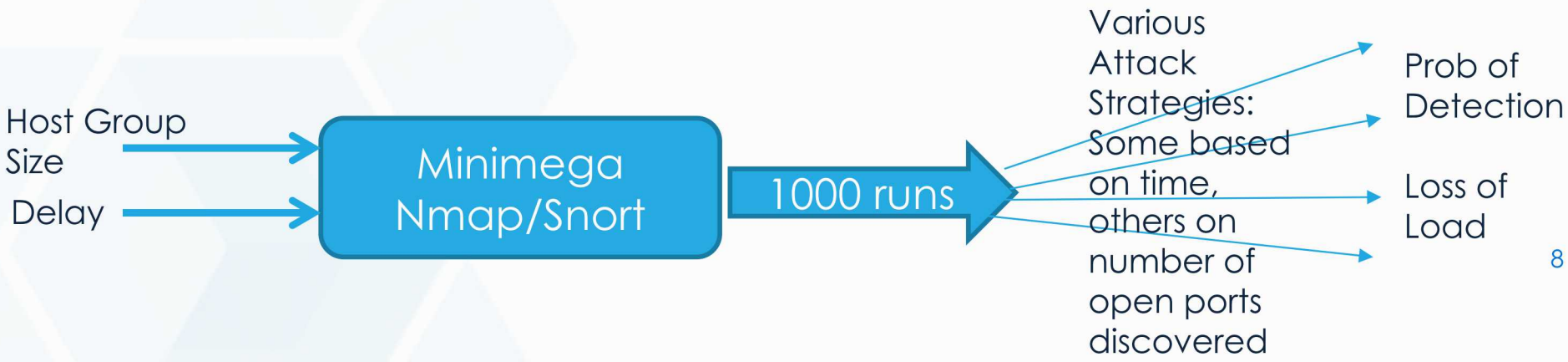


Treatment of uncertainty in scanning

- Each set of experiments in minimega involved 1000 samples at a particular setting (number of probes, delay between probes sent).
- Each of the 1000 samples was run for 200 seconds. At each second, the number of successful probes on open, closed, and filtered ports was recorded.
- These results were then post-processed using a number of attack strategies
  - One class of strategies involved the attacker just waiting to identify M open ports (M = 1, 2, 3, or 4). As soon as M ports are achieved, CRASH is deployed
  - The other class of strategies involved the attacker waiting for some time (e.g. 10, 20, 30 sec.) to deploy CRASH.
  - Each strategy then had a particular load loss, depending on the attacker/defender time race and which RTUs were hit.

Why 1000 samples? There is a large coefficient of variation (std. dev./mean) of the number of open ports found per attack strategy. For example, the T15 strategy had a CoV of 1.08, the T75 had a CoV of 0.5.

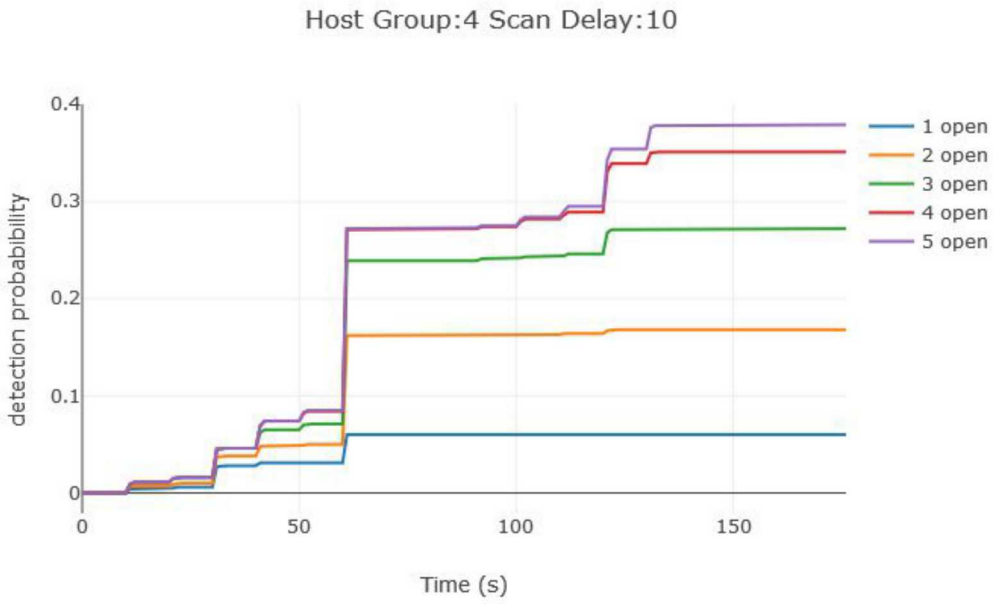
Treatment of uncertainty in scanning



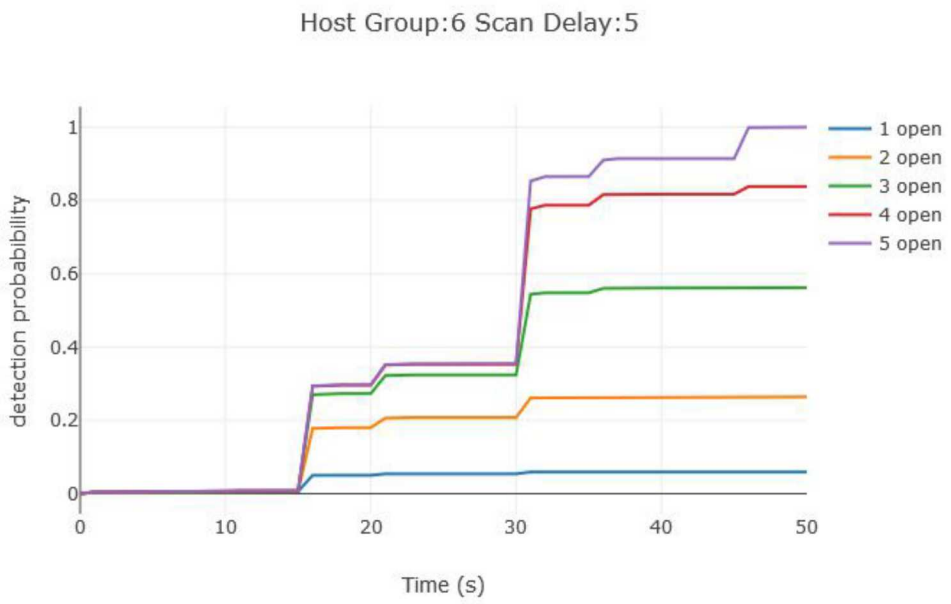
Now we can address questions such as what are the statistical differences between loss of load at 80 seconds across attack strategies?

Results: Comparing two versions of attacks

4 probes, Delay 10 sec



6 probes, Delay 5 sec



- Detection occurs much earlier when attacker runs 6 probes every 5 seconds.
- Attacker has significant probability of NOT being detected in the 4 probe, 10 second delay case.

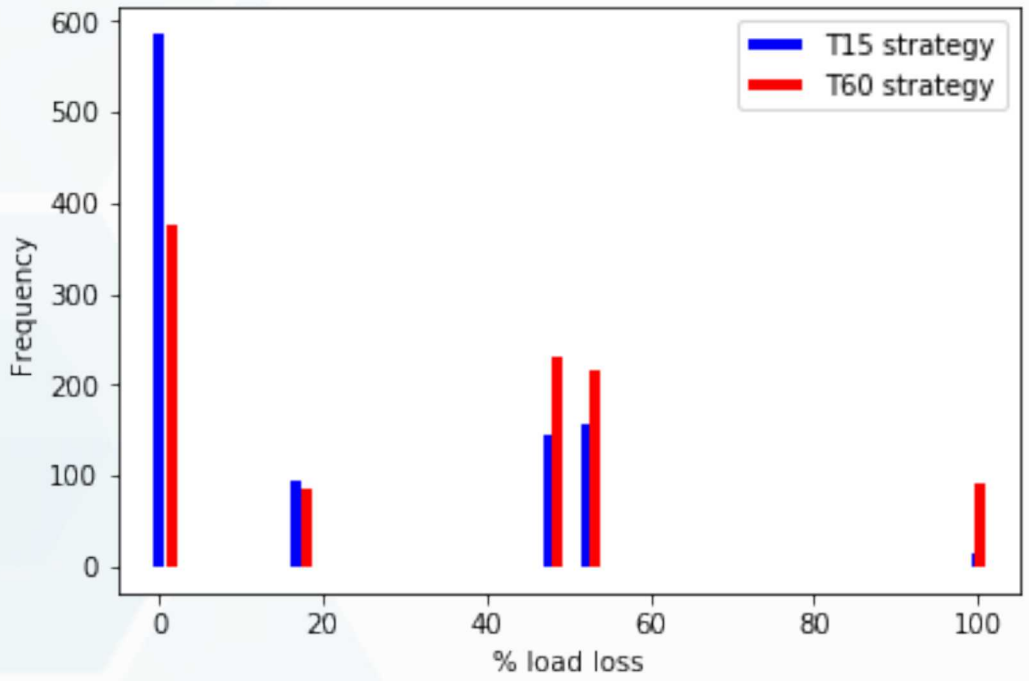
Results: Comparing two versions of attacks

4 probes, Delay 10 sec
Wait for a specified time

Next page will compare these strategies at 80 seconds.
T15 strategy: mean load loss of 18.4%
T60 strategy: mean load loss of 32.8%

Results: Zoom in on 4 probes, delay 10

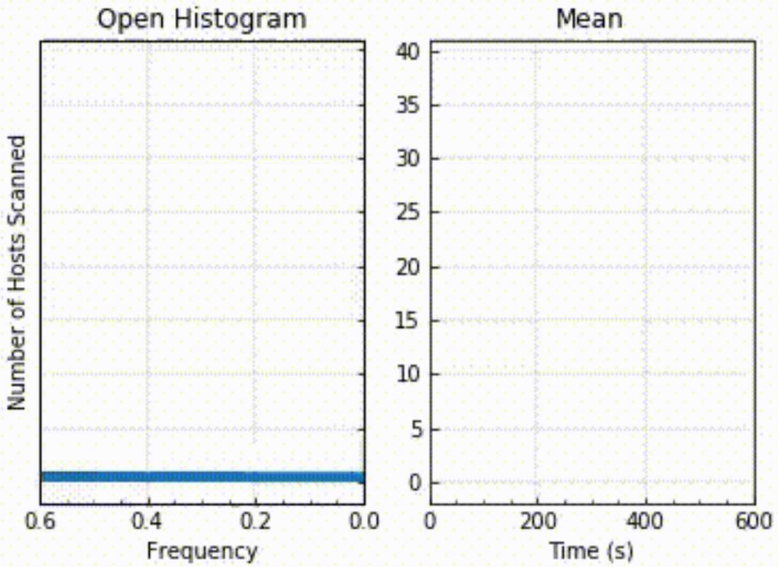
- Comparing a T15 strategy vs. a T60 strategy at 80 seconds



T15 strategy: mean load loss of 18.4%
T60 strategy: mean load loss of 32.8%

- T-test comparison for equality of mean load loss at 80 seconds using these two strategies shows that they are statistically significantly different.
- **If you only look at the mean, you don't see the differences in the distribution**

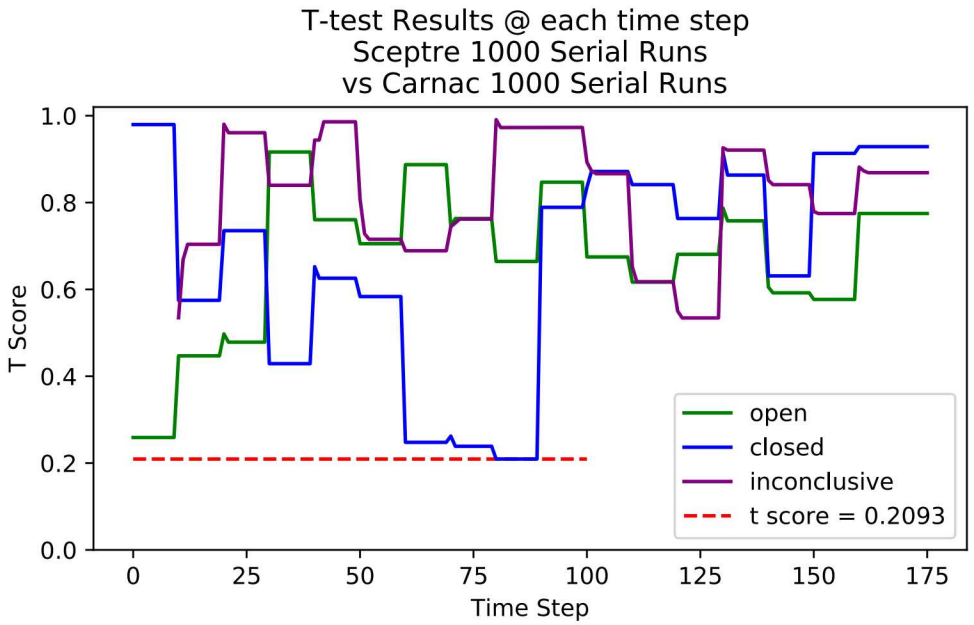
Visualization of probabilistic results over time



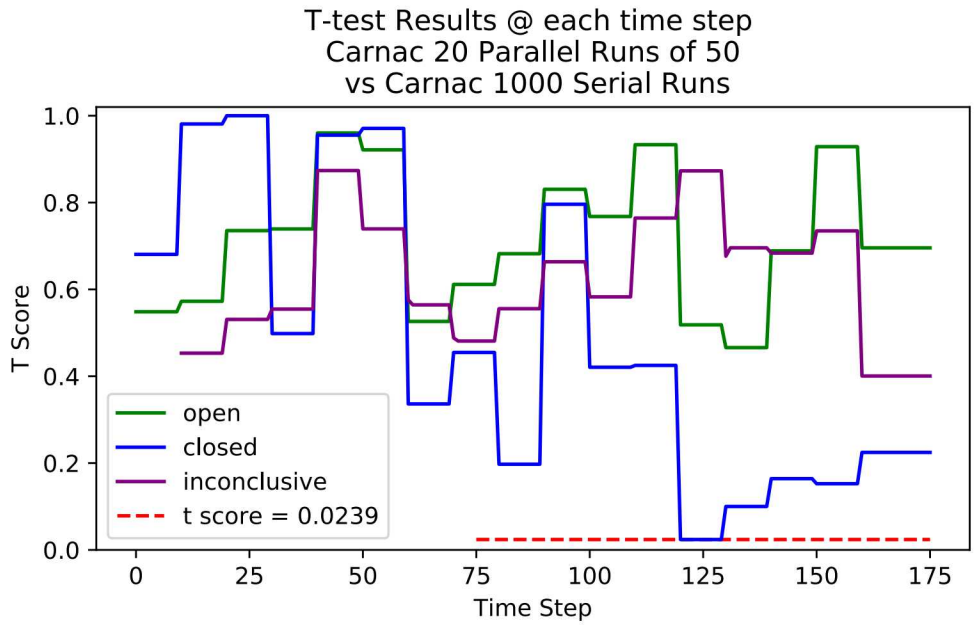
Reproducibility and Randomness

Comparison of scanning results across platforms and parallel vs. serial runs of minimega

Sceptre Serial vs Carnac Serial



Carnac Serial vs Carnac Parallel



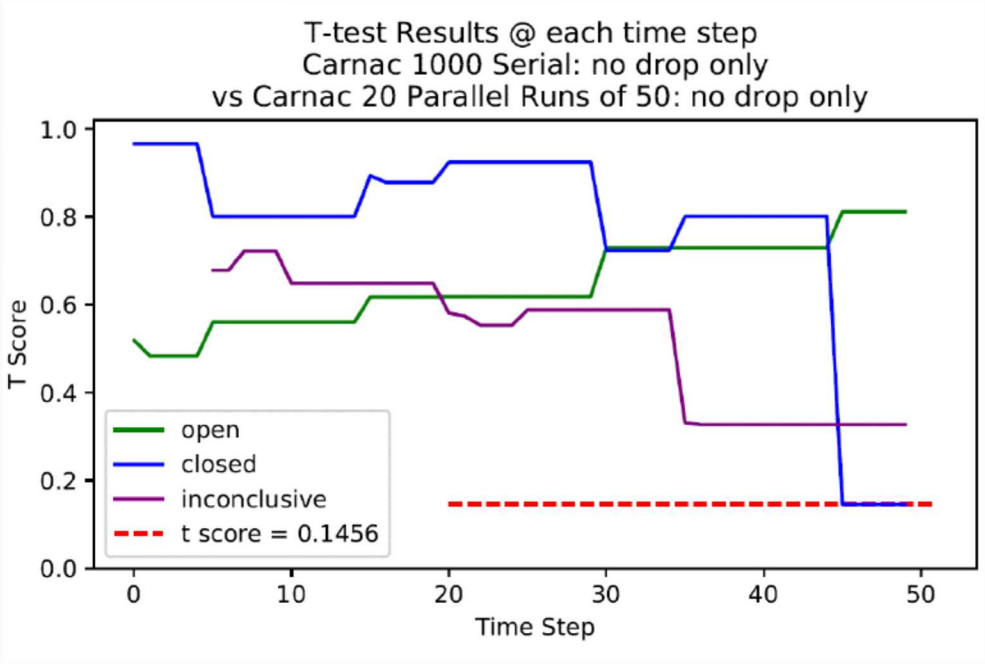
We expected more similarities in the means (higher T Score values) when comparing results across platforms or with parallel/serial implementation.
This led us to investigate randomness.

PA

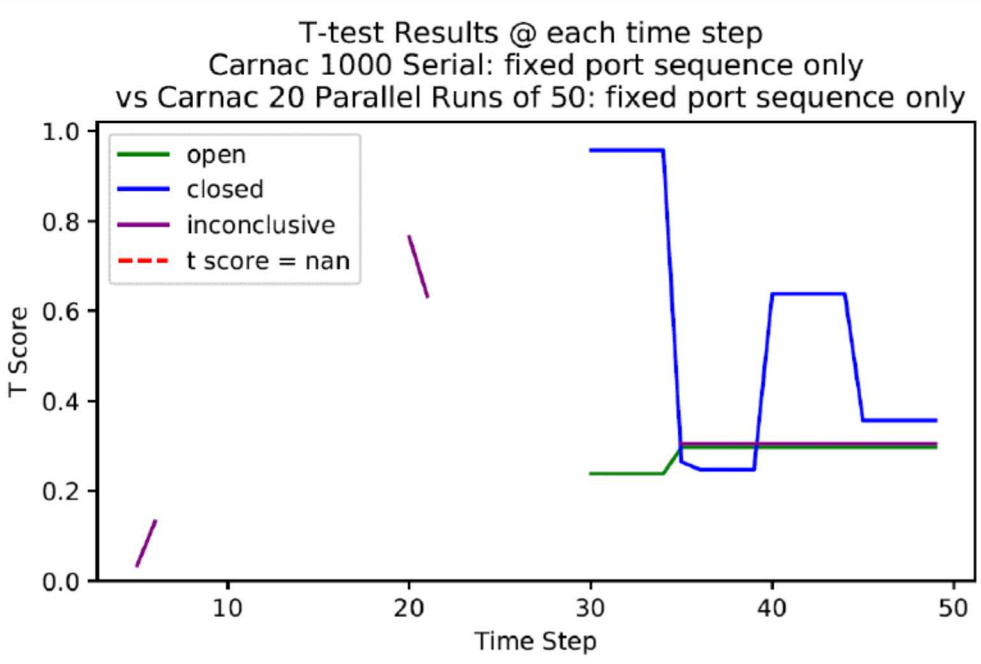
What if we remove some of the randomness?

Baseline Comparisons: remove the random aspects of the Emulytics that we control (probability of dropping a packet and random port scanning order)

No Probability of Dropping a Packet

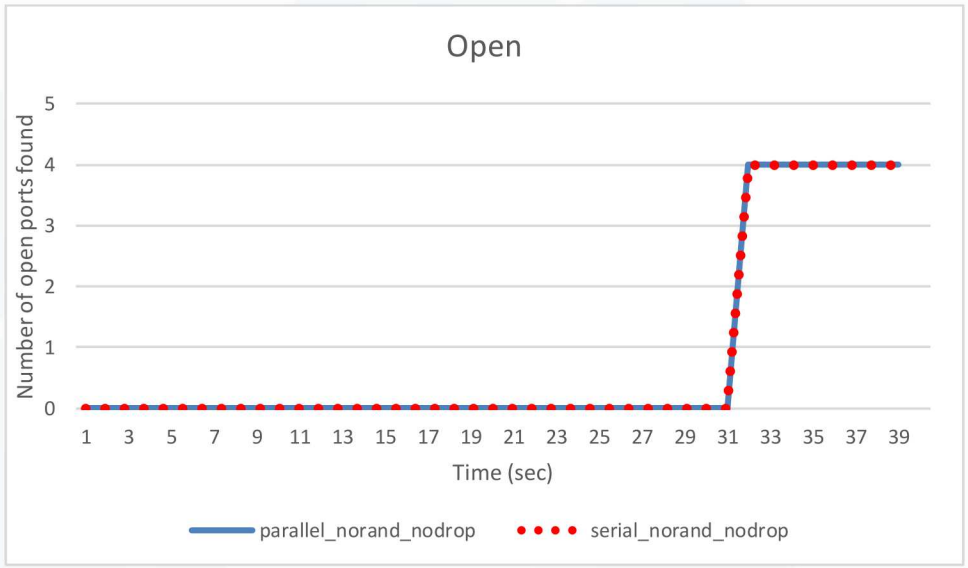


Fixed Port Order for Scanning



Deeper dive into randomness issues

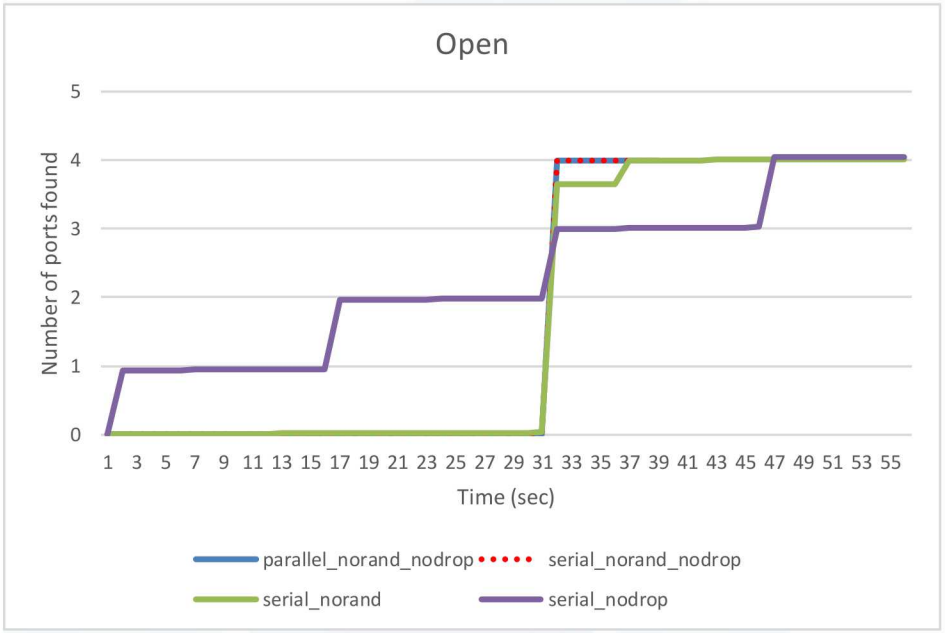
We did verify that **we get the SAME exact results across all 1000 realizations** for both serial and parallel when we have no probability of dropping a packet and a fixed port order for scanning.

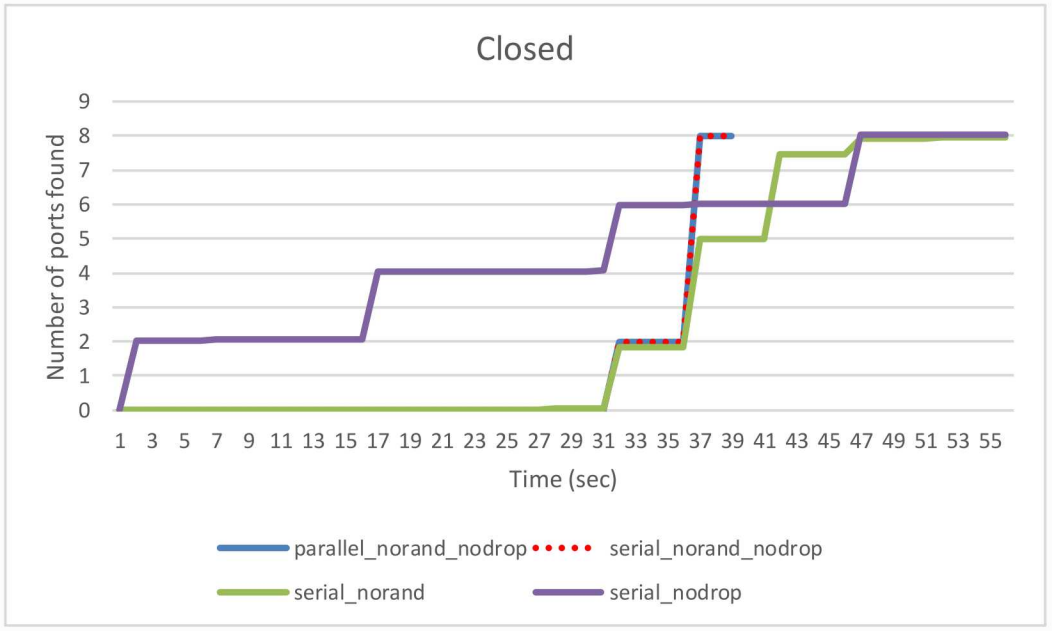


These are consistent with our understanding of the protocol and the fastest the topology can be scanned.

Deeper dive into randomness issues

Now look at effects one at a time:





Without dropped packets, ports are found earlier.
Port order has a larger effect than dropped packets.

PA

Interaction between Emulytics team and UQ team resulted in greater understanding of randomness in the emulations.

Research Thrusts

Research Thrust: Discreteness in UQ

- Polynomial Chaos Expansions (PCEs)
  - Stochastic expansions approximate the functional dependence of the output response on uncertain model parameters by expansion in a polynomial basis.
  - The polynomials used are tailored to the characterization of the uncertain variables.
  - These approaches have become very popular in the computational science community over the past two decades
  - Majority of the research is based on continuous random variables

- PCEs based on discrete polynomials:
  - Better suited to accurately represent discreteness in input variables (compared to continuous basis PCEs)
    - Expect better accuracy with fewer samples compared to continuous basis PCEs
  - Can represent full output distribution (compared to just summary statistics with MC sampling)

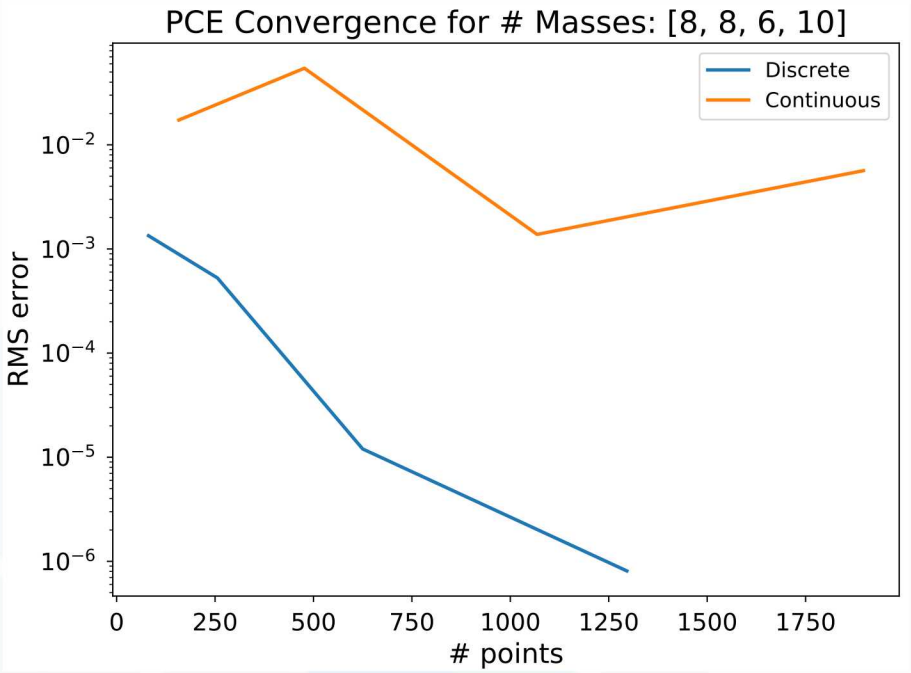
CA

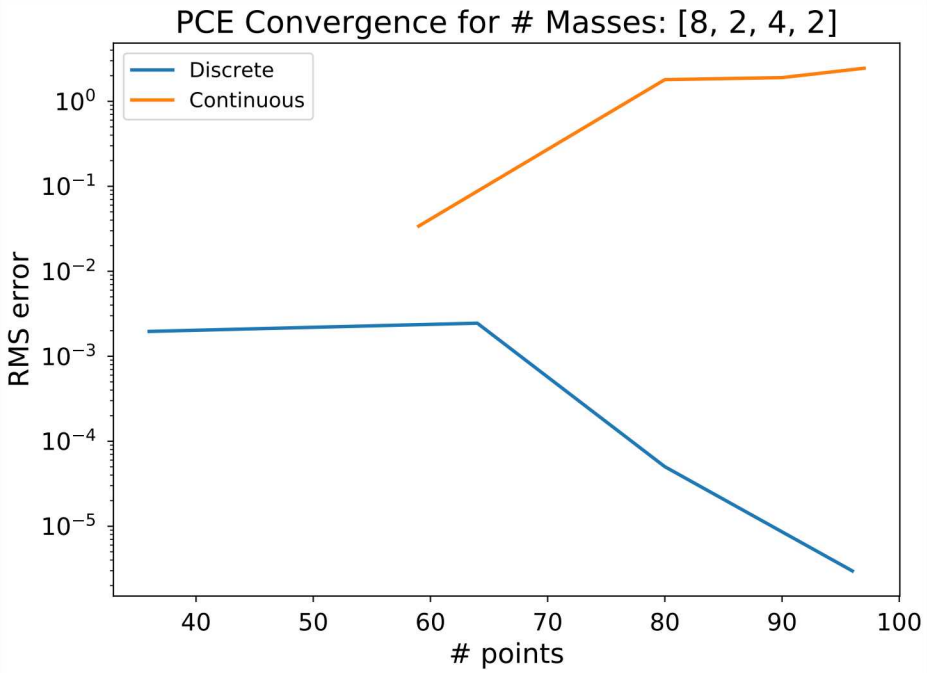
Research Thrust: Discreteness in UQ

- Pilot tests to assess feasibility and potential benefits
  - Comparing discrete and continuous PCE representations of canonical functions in terms of accuracy and number of required samples
  - Analyzing cost (number of samples) as a function of
    - Number of uncertain variables
    - Number of levels for each discrete variable
    - Nonlinearity of the approximated function
  - In collaboration with John Jakeman and Cosmin Safta through the FASTMath SciDAC institute
- On next slide, the discrete PCE is calculated in terms of custom polynomials with coefficients obtained through regression on Leja samples
- On the next slide, the continuous Legendre-Uniform PCE is obtained through regression on points randomly sampled in discrete input space

CA

Discrete PCE outperforms continuous, especially when some dimensions have few masses.





Input space of size 8 x 8 x 6 x10=3840

Input space of size 8 x 2 x 4 x 2=128

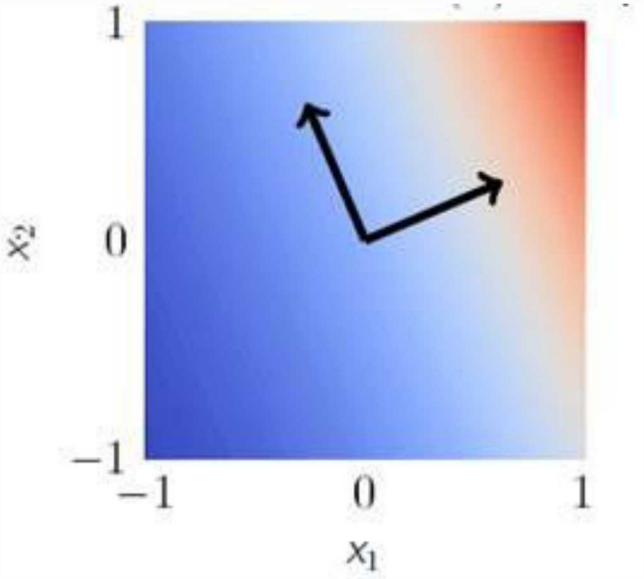
- Approximation of Genz Gauss Peak Function
- 4D discrete variables with uniform probability masses
  - Number of masses in each dimension randomly picked
  - Locations randomly picked in each dimension

Research Thrust: Multifidelity Modeling

- Gianluca Geraci will present slides on this topic
- Minimega emulator is used as the high-fidelity model, NS-3 network simulator as the low-fidelity model
- Showing progression of multifidelity methods with a series of case studies of increasing complexity:
  - QoI is the response (requests/second) for http traffic
  - 1 client/1server
  - 1 source/1 destination but four routers in between. The routing is fixed.
  - Same as above but the routing paths are given different costs to demonstrate the effects of changing routes (e.g. changing topology)
  - Bandwidth rates, delays, and number and size of packets are input variables.

Dimension Reduction

- Two approaches:
  - Explicit aggregation of nodes (100 nodes aggregated to one which has a similar behavior as the 100 in terms of traffic, loads)
  - Formal mathematical approaches
    - We are starting to get large, rich data sets (e.g. the closed/open/filtered states over time, or all the power states over time).
    - 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
    - Efficiency for UQ



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

FY20 Plan: Scalability and Tail Estimation

- Research Thrusts
  - Multifidelity UQ:
    - Can we scale from 6 nodes, 14 uncertainties to a hundred nodes, hundreds of uncertainties?
    - Can we include discrete choices (e.g. topology routing) within MF UQ at scale?
    - Develop multifidelity approaches for tail estimation. CA
  - Discrete polynomials:
    - What is the limit for this approach? What are advantages and disadvantages compared with plain MC? CA
  - Dimension reduction: Take rich state information from the scanning/detection/power systems state output and start with PCA.
  - Continue to support exemplar uncertainty and sensitivity analysis studies
  - Validation studies
  - SECUREtk development

Publication Plan

Publication	Milestone Date
International Conference on Uncertainty Quantification in Computational Sciences and Engineering	19Q3
12th USENIX Workshop on Cyber Security Experimentation and Test (CSET)	19Q4
INFORMS 2019 Annual Meeting	20Q1
Multifidelity approaches for Emulytics models: SIAM/ASA Journal on Uncertainty Quantification	20Q4
13th USENIX Workshop on Cyber Security Experimentation and Test (CSET)	20Q4
SIAM Computational Science and Engineering	21Q2
14th USENIX Workshop on Cyber Security Experimentation and Test (CSET)	21Q4
Experimental Design/Dimension reduction for Emulytics models: Journal of Network and Computer Applications	21Q4

LDRD

Laboratory Directed Research and Development

Multifidelity UQ for network applications: Lessons learned and perspectives

Gianluca Geraci, Jonathan Crussell, Laura Swiler, Bert Debusschere and Erin Acquesta

SECURE LDRD Grand Challenge
External Advisory Board
October 29th, 2019

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.

Multifidelity Uncertainty Quantification
(a recap from the previous EAB)

Uncertainty Quantification
Forward Propagation – Why Sampling Methods?

(Forward) **Uncertainty Quantification:** propagate the uncertainty parameters through the computer codes in order to quantify their effects on the Quantity of Interest (QoIs)

UQ context for SECURE at a glance:

- High-dimensionality, non-linearity and bifurcations/discontinuities
- Large set of modeling choices available (network topology, operative conditions, etc.)

Natural candidate for UQ:

- **Sampling**-based (MC-like) approaches because they are **non-intrusive**, **robust** and **flexible**...
- **Drawback:** Slow convergence $\mathcal{O}(N^{-1/2}) \rightarrow$ many realizations to build reliable statistics

Goal of Multifidelity UQ:

Reducing the computational cost of obtaining MC reliable statistics by combining several models

Pivotal idea:

- Simplified (**low-fidelity**) models are **inaccurate** but **computationally inexpensive**
  $\Rightarrow$ **low-variance** estimates
- **High-fidelity** models are **costly**, but **accurate**
  $\Rightarrow$ **low-bias** estimates

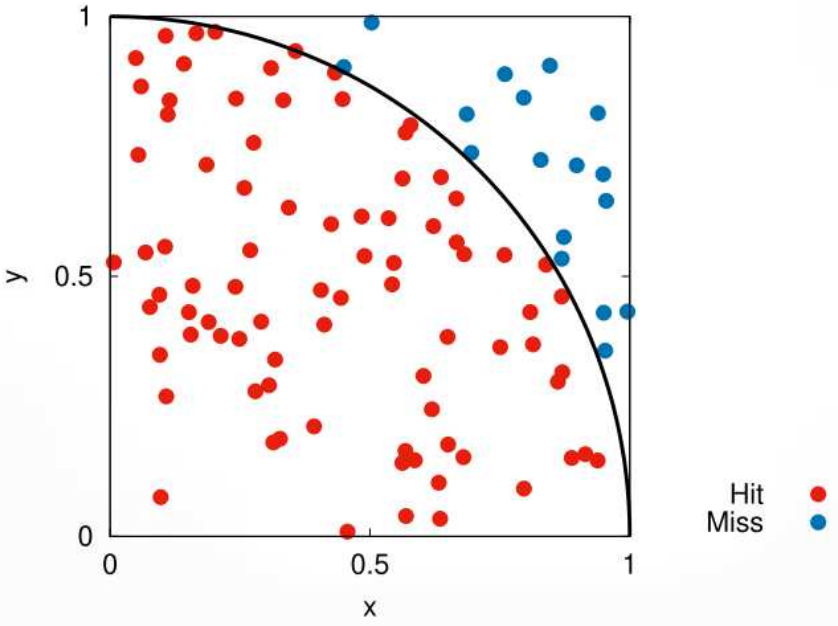
Sampling methods are **complementary with respect to the (discrete) surrogates** approaches

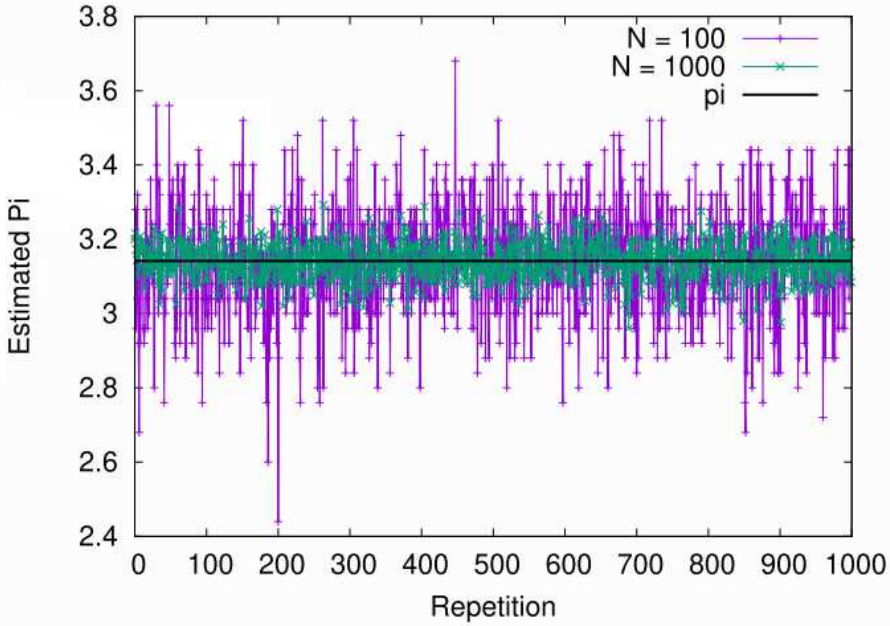
How does a sampling method work?

Let consider a **random variable** Q, we want to compute **its expected value** $\mathbb{E}[Q]$ (or high-order moments):

$$\hat{Q}_N^{\mathrm{MC}} = \frac{1}{N} \sum_{i=1}^{N} Q^{(i)}$$

Let's use MC to compute the value $\pi \propto \dfrac{\#\mathrm{Hit}}{N}$

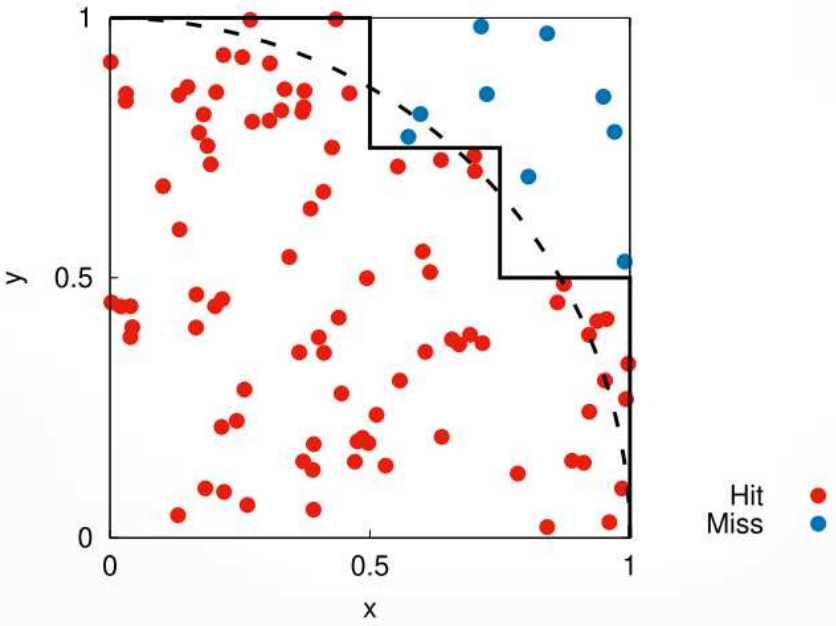


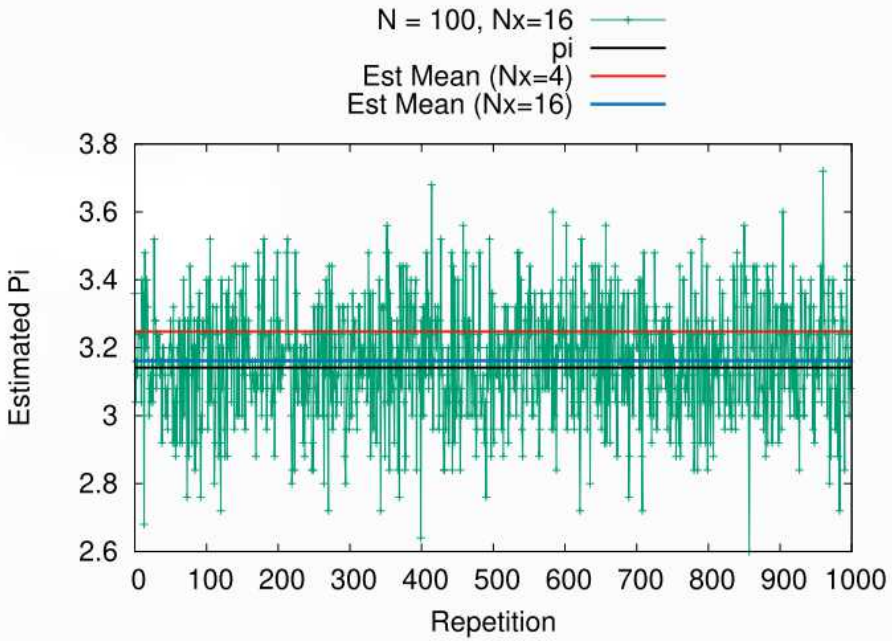


Numerical problems **cannot be resolved with infinite accuracy** (discretization error), the MC estimator for a specific **fidelity** M**th** level

$$\hat{Q}^{MC}_{\mathsf{M},N} \overset{\text{def}}{=} \frac{1}{N}\sum_{i=1}^{N} Q_{\mathsf{M}}{}^{(i)}$$

Let's use MC to compute the value $\pi \propto \dfrac{\#\text{Hit}}{N}$





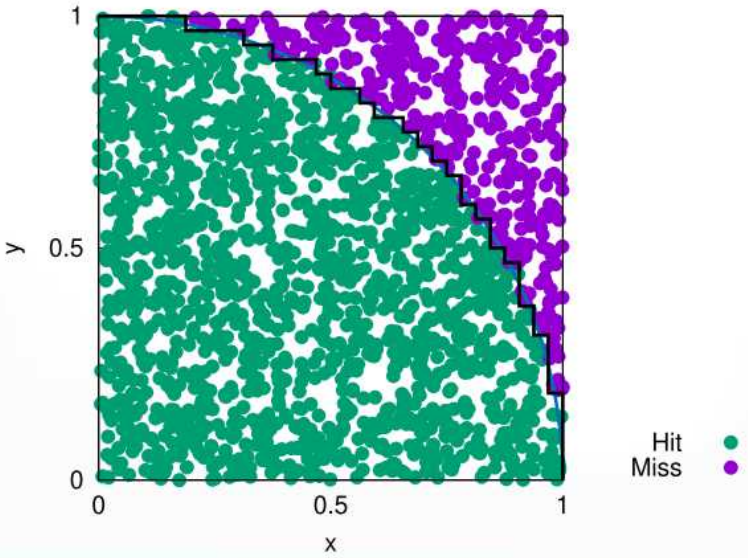
Two sources of error in the **Mean Square Error**:

$$\mathbb{E}\left[(\hat{Q}^{MC}_{M,N} - \mathbb{E}[Q])^2\right] = \frac{\mathbb{V}\mathbf{ar}(\mathbf{Q})}{\mathbf{N}} + (\mathbb{E}[\mathbf{Q_M} - \mathbf{Q}])^2$$

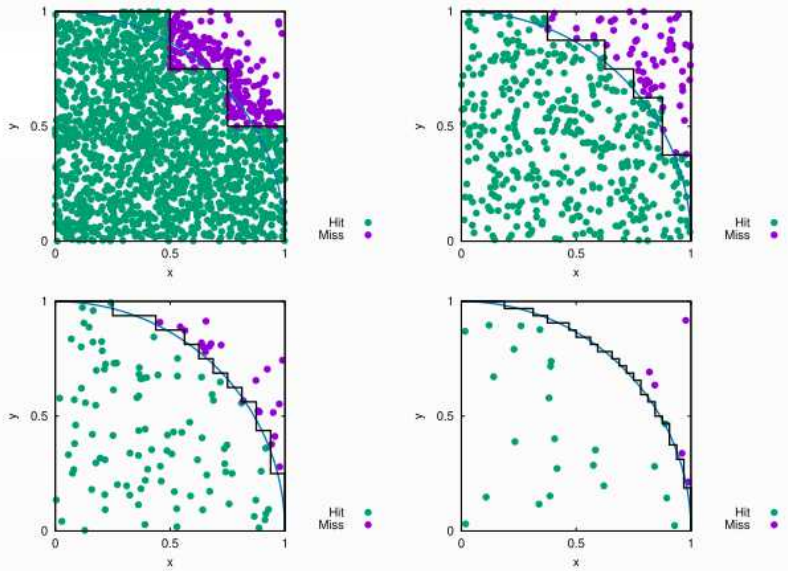
Pivotal idea:

- **High-fidelity** models are **costly**, but **accurate**
  - **low-bias** estimates
- Simplified (**low-fidelity**) models are **inaccurate** but **cheap**
  - **low-variance** estimates

Single Fidelity



Multi Fidelity



Multifidelity Estimator: How does it work?

Optimal Control Variate

M low-fidelity models with known expected value

What do we need?

- HF model, *i.e.* the model for which you want to compute the statistics
- (a set of) LF model(s), *i.e.* they are not required to predict the HF response but only to be correlated

Optimal Control Variate

M low-fidelity models with known expected value

What do we need?

- HF model, *i.e.* the model for which you want to compute the statistics
- (a set of) LF model(s), *i.e.* they are not required to predict the HF response but only to be correlated

How do we build the MF estimator?

1. We take a MC estimator for the HF model, $\hat{\mathbf{Q}}$

Optimal Control Variate

M low-fidelity models with known expected value

What do we need?

- HF model, *i.e.* the model for which you want to compute the statistics
- (a set of) LF model(s), *i.e.* they are not required to predict the HF response but only to be correlated

How do we build the MF estimator?

1. We take a MC estimator for the HF model, $\hat{\mathbf{Q}}$
2. We add a weighted sum of **unbiased terms**, $\sum_{\mathbf{i=1}}^{\mathbf{M}} \alpha_{\mathbf{i}} \left(\hat{\mathbf{Q}}_{\mathbf{i}} - \hat{\mu}_{\mathbf{i}} \right)$ where $\hat{\mu}_i$ is an approximation to the expected value of the ith LF model

Optimal Control Variate

M low-fidelity models with known expected value

What do we need?

- HF model, *i.e.* the model for which you want to compute the statistics
- (a set of) LF model(s), *i.e.* they are not required to predict the HF response but only to be correlated

How do we build the MF estimator?

1. We take a MC estimator for the HF model, $\hat{\mathbf{Q}}$
2. We add a weighted sum of **unbiased terms**, $\sum_{\mathbf{i=1}}^{\mathbf{M}} \alpha_{\mathbf{i}} \left(\hat{\mathbf{Q}}_{\mathbf{i}} - \hat{\mu}_{\mathbf{i}} \right)$ where $\hat{\mu}_i$ is an approximation to the expected value of the ith LF model
3. We consider N_i LF evaluations: $N_i = \lceil r_i N \rceil$ for each model

Optimal Control Variate

M low-fidelity models with known expected value

What do we need?

- HF model, *i.e.* the model for which you want to compute the statistics
- (a set of) LF model(s), *i.e.* they are not required to predict the HF response but only to be correlated

How do we build the MF estimator?

1. We take a MC estimator for the HF model, $\hat{\mathbf{Q}}$
2. We add a weighted sum of **unbiased terms**, $\sum_{\mathbf{i=1}}^{\mathbf{M}} \alpha_{\mathbf{i}} \left(\hat{\mathbf{Q}}_{\mathbf{i}} - \hat{\mu}_{\mathbf{i}} \right)$ where $\hat{\mu}_i$ is an approximation to the expected value of the ith LF model
3. We consider N_i LF evaluations: $N_i = \lceil r_i N \rceil$ for each model
4. We solve for the optimal weights $\alpha_{\mathbf{i}}$ (and the optimal number of LF evaluations N_i)

Optimal Control Variate

M low-fidelity models with known expected value

What do we need?

- HF model, *i.e.* the model for which you want to compute the statistics
- (a set of) LF model(s), *i.e.* they are not required to predict the HF response but only to be correlated

How do we build the MF estimator?

1. We take a MC estimator for the HF model, $\hat{\mathbf{Q}}$
2. We add a weighted sum of **unbiased terms**, $\sum_{\mathbf{i=1}}^{\mathbf{M}} \alpha_{\mathbf{i}} \left(\hat{\mathbf{Q}}_{\mathbf{i}} - \hat{\mu}_{\mathbf{i}} \right)$ where $\hat{\mu}_i$ is an approximation to the expected value of the ith LF model
3. We consider N_i LF evaluations: $N_i = \lceil r_i N \rceil$ for each model
4. We solve for the optimal weights $\alpha_{\mathbf{i}}$ (and the optimal number of LF evaluations N_i)

What do we obtain?

$$\hat{Q}^{\mathrm{ACV}} = \hat{\mathbf{Q}} + \sum_{\mathbf{i=1}}^{\mathbf{M}} \alpha_{\mathbf{i}} \left(\hat{\mathbf{Q}}_{\mathbf{i}} - \hat{\mu}_{\mathbf{i}} \right)$$

$$\mathbb{V}ar\left(\hat{Q}^{\mathrm{ACV}} \right) = \mathbb{V}\mathbf{ar}\left(\hat{\mathbf{Q}} \right) \left(\mathbf{1} - \mathbf{R}^{\mathbf{2}}_{\mathrm{ACV}} \right).$$

How does the variance reduction term look like?

- For a single low-fidelity model: $R^2_{ACV-1} = \frac{\mathbf{r_1} - 1}{\mathbf{r_1}} \rho_{\mathbf{1}}^{\mathbf{2}}$, where $\mathbf{r_1} = \sqrt{\frac{\mathcal{C}_{\mathbf{HF}}}{\mathcal{C}_{\mathbf{LF}}} \frac{\rho_{\mathbf{1}}^{\mathbf{2}}}{1 - \rho_{\mathbf{1}}^{\mathbf{2}}}}$
- (Pearson's) correlation coefficient: $\rho_{\mathbf{1}}$
- Computational cost ratio: $w = \frac{\mathcal{C}_{\mathbf{HF}}}{\mathcal{C}_{\mathbf{LF}}}$

$$C_{MF}^{tot} = C_{MC}^{tot}\ \Theta\left(w = \frac{C_{HF}}{C_{LF}}, \rho^2\right)$$

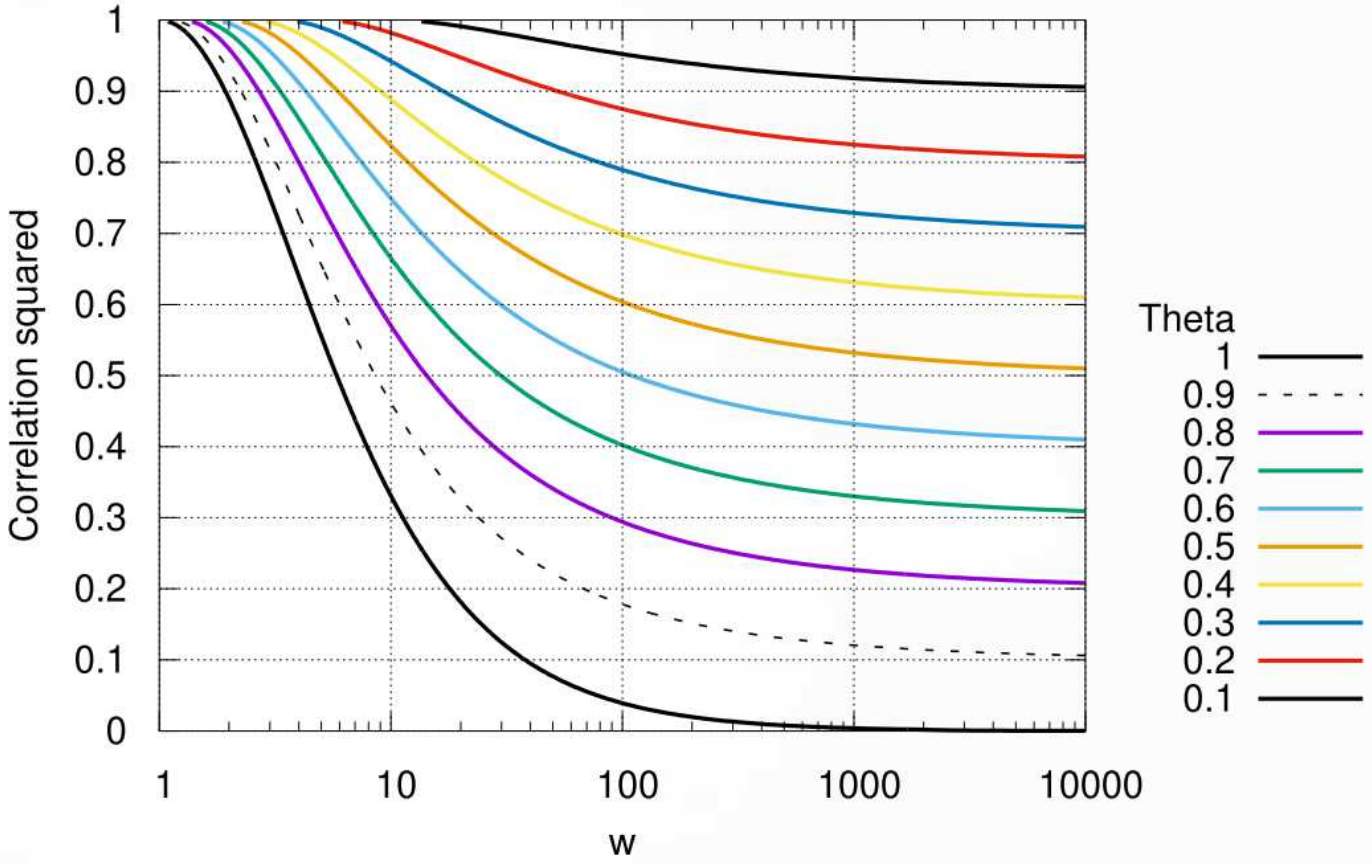
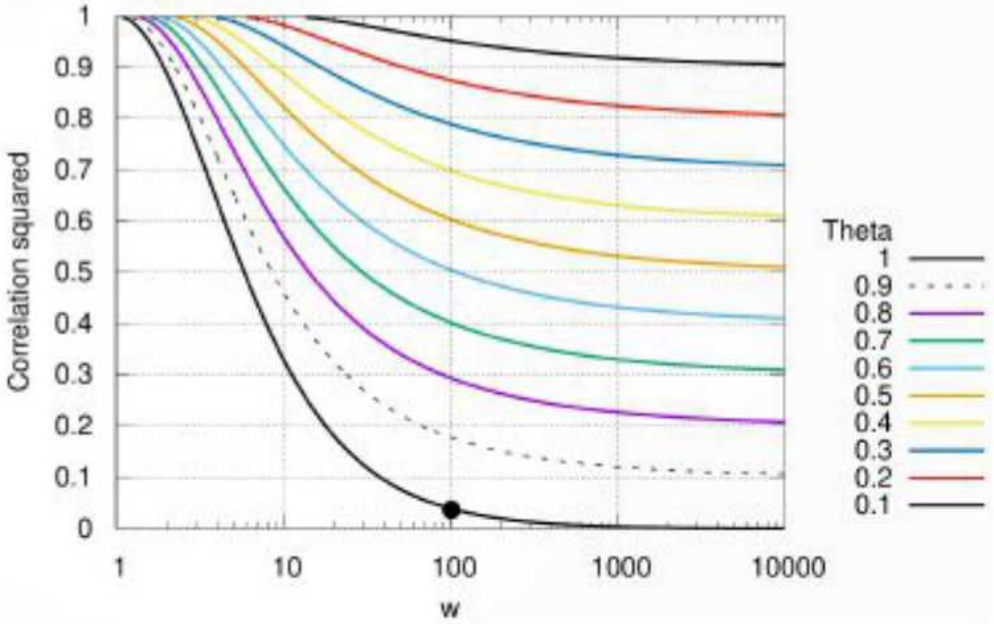
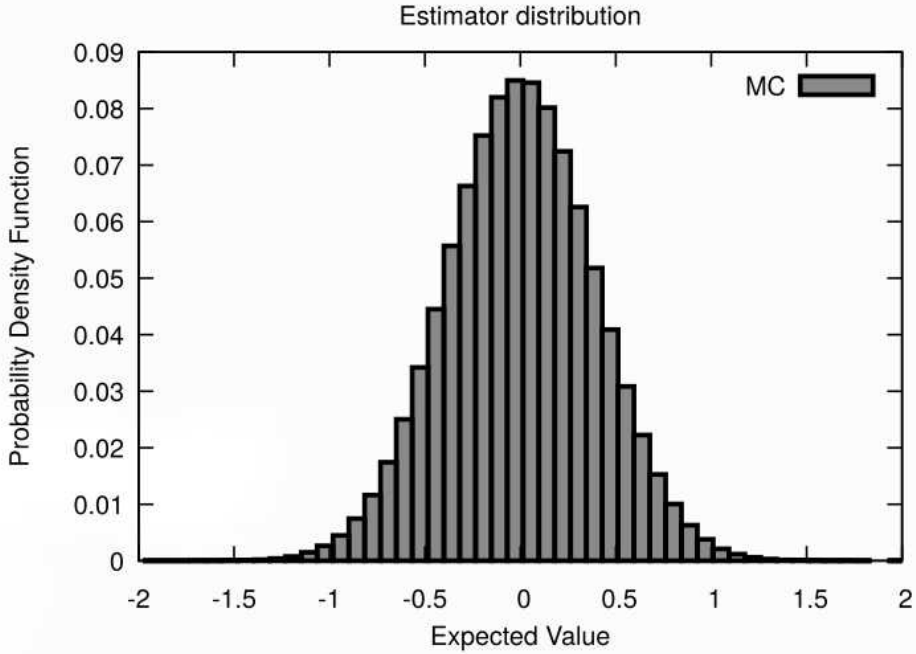


FIGURE: MF normalized total cost w.r.t. to a MC with same estimator variance.

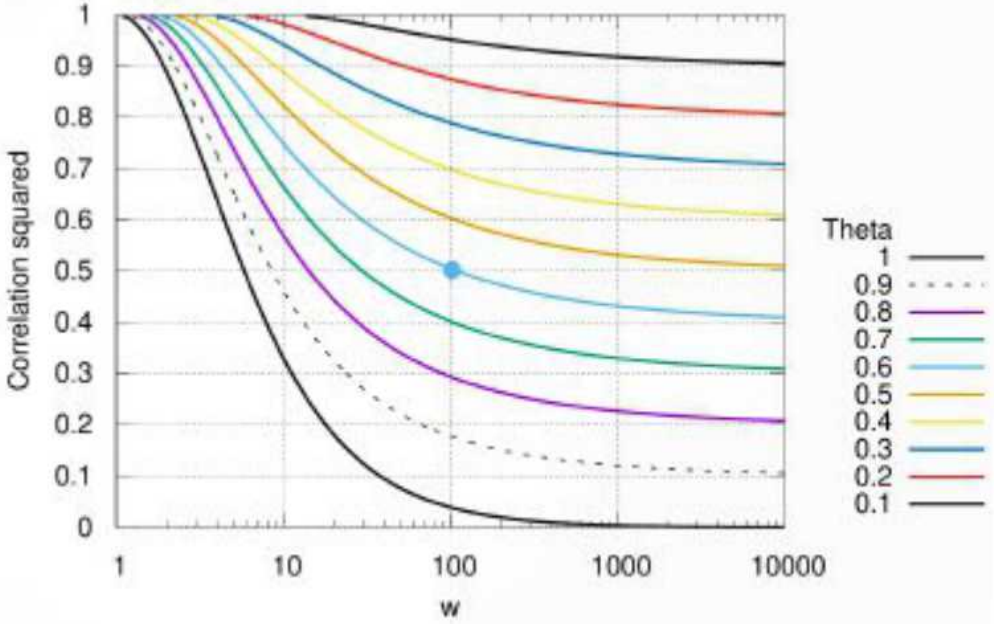
Multifidelity Estimator

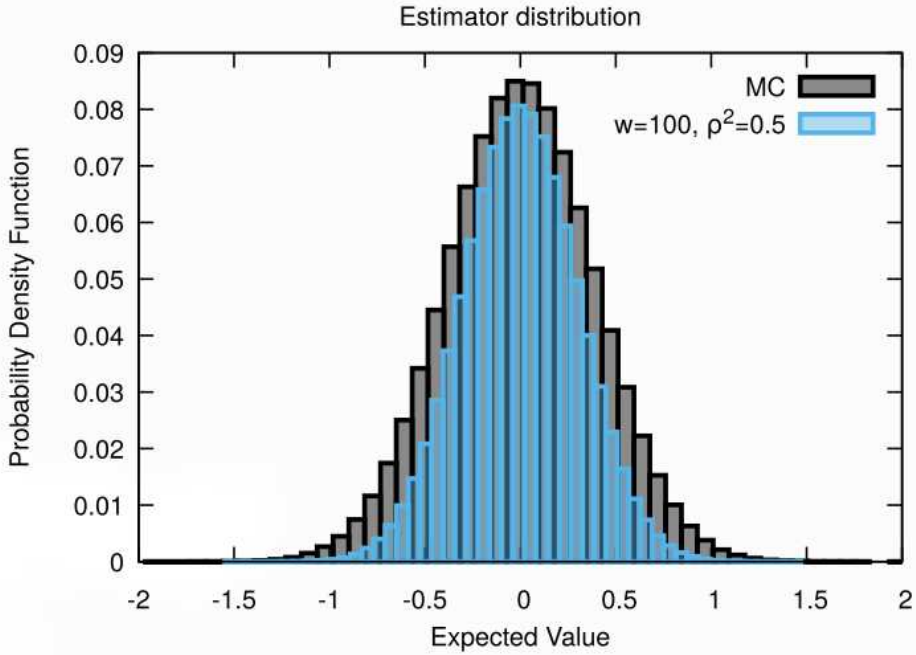
How does it compare with MC?





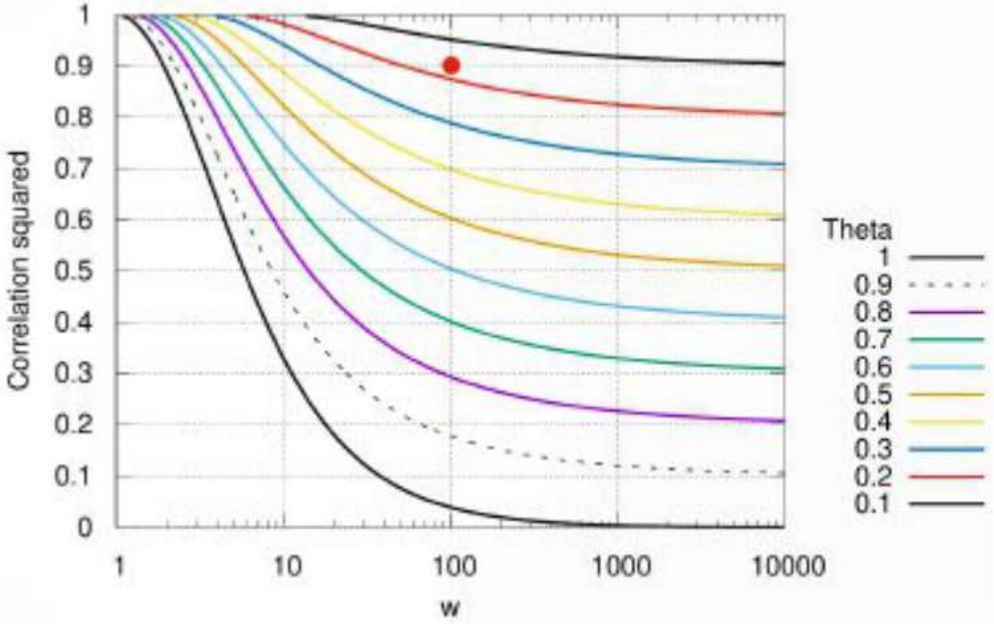
Multifidelity Estimator
How does it compare with MC?

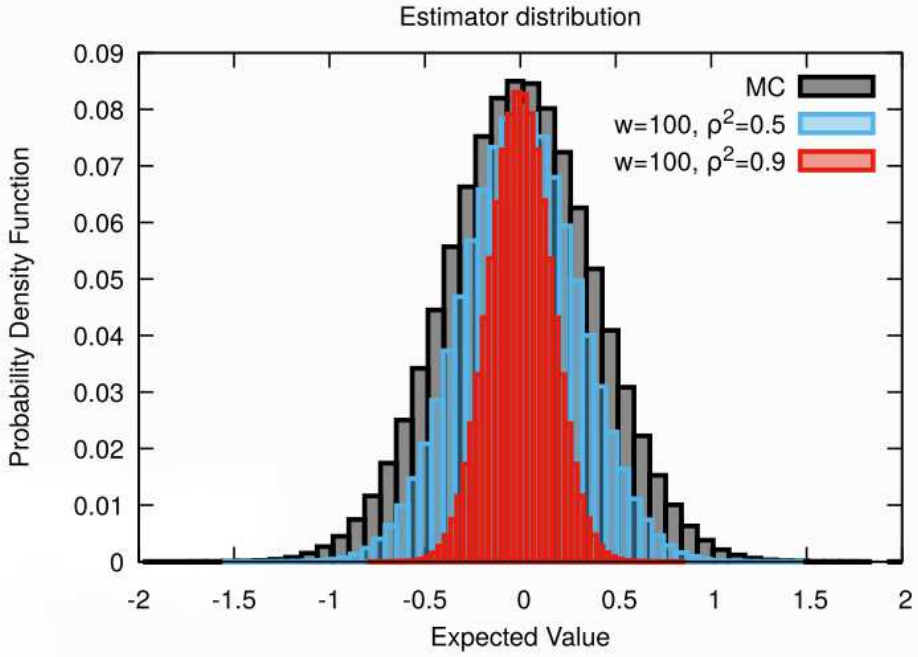




Variance Multifidelity $= 0.6 \times$ **Variance MC**

Multifidelity Estimator
How does it compare with MC?





Variance Multifidelity $= 0.17 \times$ **Variance MC**

Numerical Experiments

minimega

- Tool to launch, manage and instrument virtual machines and networks
- It can run on your laptop or distributed across a cluster
- Scriptable API for automated experimentation
- Open source GNU GPLv3-licensed, publicly available and active project
- Integrate real hardware or humans with virtual experiments

Network emulation

- Experiments run in real time on virtualized hardware
- Initialization phase to launch VMs (OSes, applications, *etc.*)
- Virtual hardware introduces artifacts (*i.e.* stochasticity) from shared resources and nested functionalities
- Running real software captures real system behaviors
- Allows for heterogeneous OSes
- Flexible with respect to unknown software (does not require source code)

ns-3

- ns-3 is a **discrete event** simulator for IP and non-IP addresses
- Software written in C++ with bindings available for Python
- GNU GPLv2-licensed
- Possible to construct simulations from **reusable components** to configure nodes, topologies and applications

Discrete-event simulation

- Time evolves from event to event
- A single-threaded event list is executed
- Events are scheduled to occur at specific virtual/simulation time
- Events can generate additional events
- Simulation ends when a specific time is reached or there are no more events

Few comments on the State-of-the-art:

- Uncertainty Quantification is a **relatively new concept in network applications**
- Multifidelity UQ is a **new concept in the UQ realm**
- MF UQ for Emulytics is going to unveil **challenges that cannot be entirely anticipated**

Progression of test cases with increasing complexity:

- 1 Client - 1 Server example
  - Is the concept of low-fidelity applicable in computer network applications?
  - Is `ns-3` a viable way of constructing such low-fidelity models?
  - How much difficult is it to obtain a correlated low-fidelity model?
- 4 routers case with fixed costs
  - Can we still apply MF UQ for a more complex (fixed) topology?
  - How much difficult is it to obtain a correlated low-fidelity model for a more realistic scenario, *i.e.* higher number of uncertainty parameters?
- 4 routers case with varying costs (*i.e.* varying topology)
  - How good is *ns*-3 in capturing the response if the topology changes?

What can UQ provide today?

- Our experience with a variety of applications in Computational Science suggests to us that performing UQ studies can help obtaining beneficial information starting from the **verification process** throughout the entire **system validation**

Test #1: 1 Client - 1 Server (`minimega-ns-3`)

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

- **Quantity of interest:** Number of requests/s
- `minimega` – HF: 100 Requests (average over 10 repetitions)
- `ns-3` – LF: 10 Requests (`Delay` $50ms$)
- `ns-3` – LF*: 1 Requests (`Delay` $5ms$)

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

Table: Normalized Cost

We assume **serial execution for the LF model**, however we might easily increase the efficiency of LF (`ns-3`) by running multiple concurrent evaluations

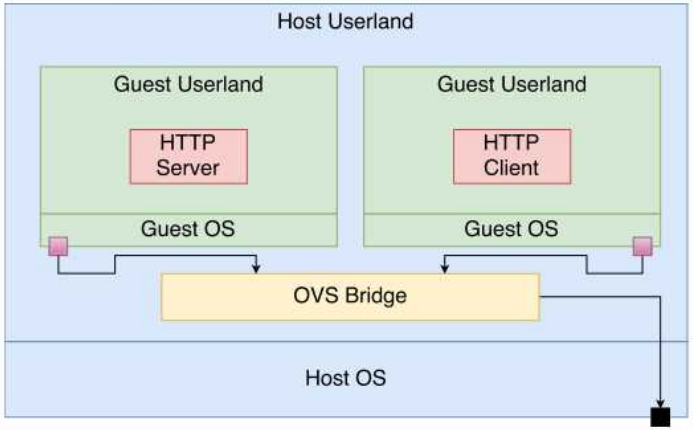


Figure: Network Configuration

First minimega–ns-3 demonstration
Estimator Standard Deviation

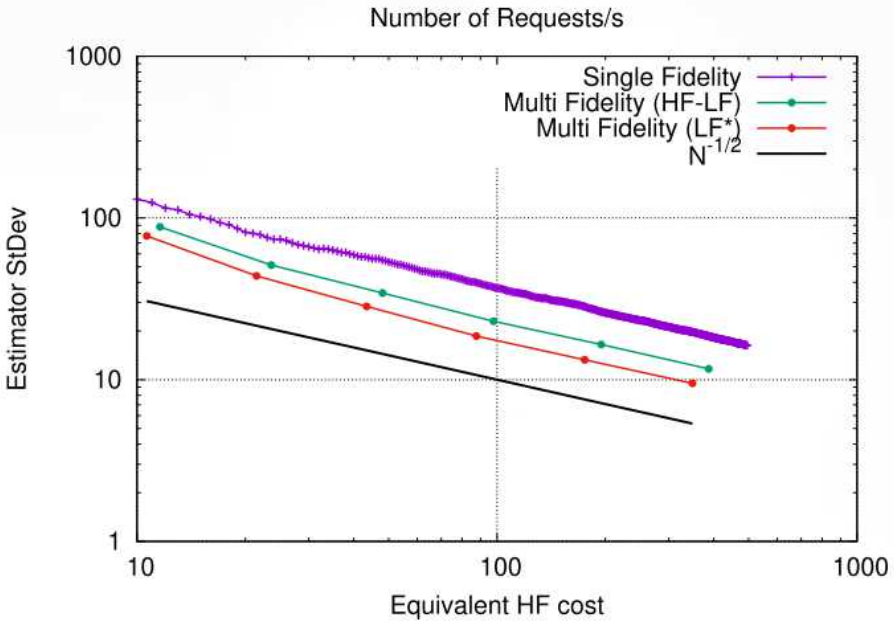


FIGURE: Exp. Value StDev

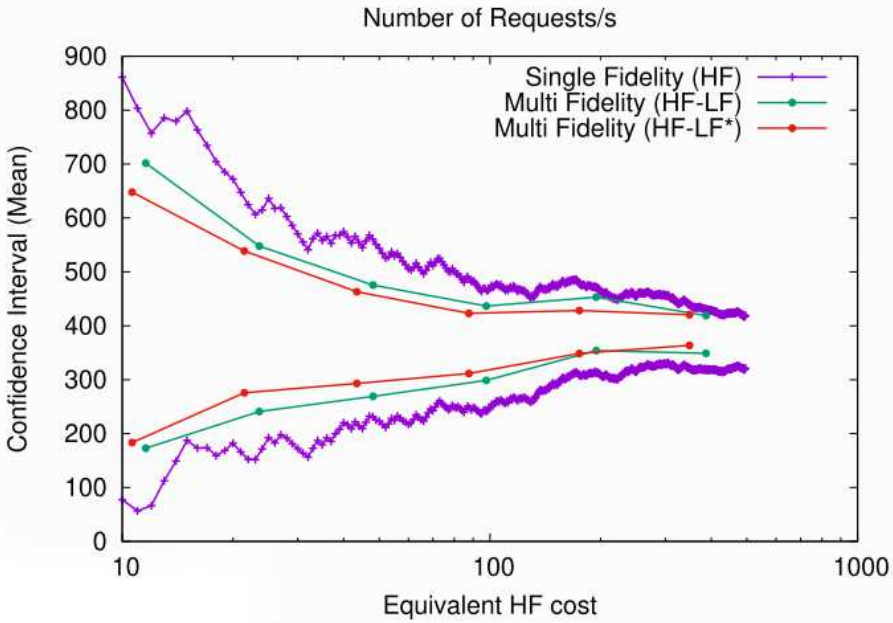


FIGURE: Exp. Value Confidence Interval

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

Test #2: 4 routers case with fixed costs (`minimega-ns-3`)

Network Configuration

- Source and Destination separated by 4 (non-aligned) routers
- 2000 Requests

Uncertain Parameters (7 parameters)

- `DataRate` $\sim \mathcal{U}(5, 500)$Mbps
- `Delay` **fixed** to $2ms$

Fidelity definition

- **Quantity of interest:** Number of Requests/s
- HF (`minimega`): `ResponseSize` $100KB$ (average over 5 iterations)
- LF (`ns-3`): `ResponseSize` $50B$ and 10 Requests

	C
HF	1
LF	2.45E-4

Table: Normalized Cost

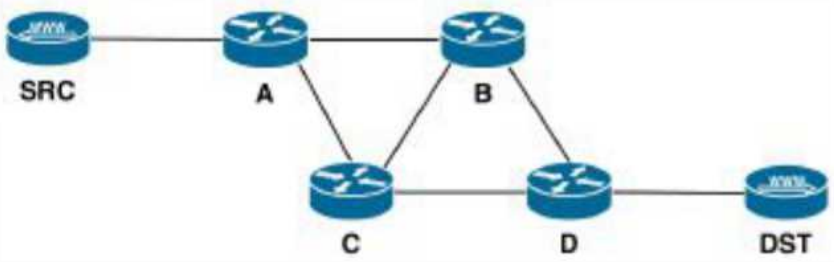


Figure: Network Configuration

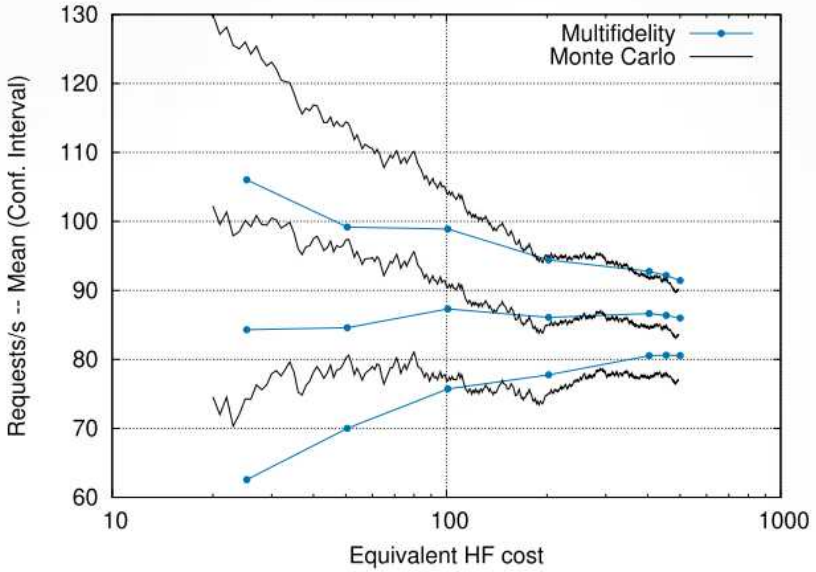


FIGURE: Estimated means and CIs.

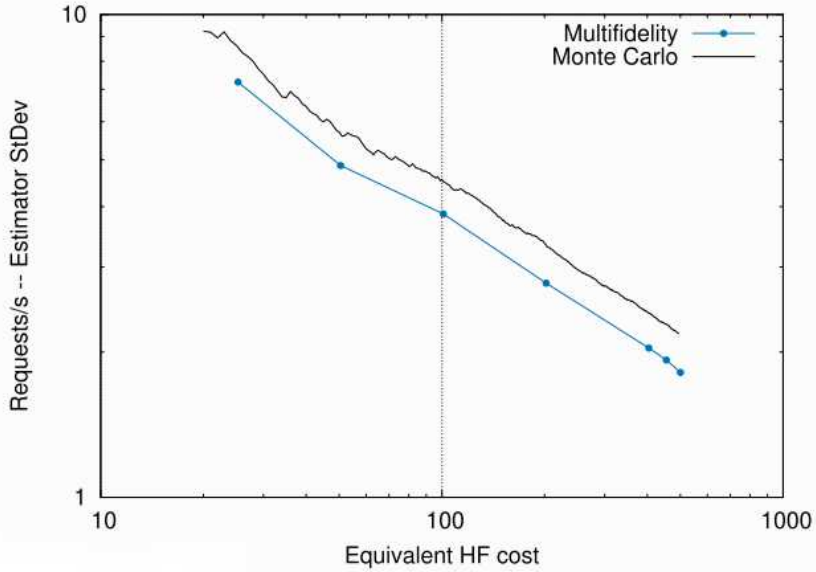


FIGURE: Standard deviation.

Notes:

- 30% variance reduction
- Correlation ~ 0.56 and $r = 43$

Test #3: 4 routers case with uncertain costs

(Can we use UQ principles to understand better `minimega`'s response?)

Network Configuration

- Source and Destination separated by 4 (non-aligned) routers – This case has 5 edges
- 2000 Requests

Uncertain Parameters (14 parameters: 8 rates and 6 costs)

- `DataRate` $\sim \mathcal{U}(5, 500)$Mbps
- `Cost` $\sim \mathcal{U}(1, 4)$ (cost for A–D is $\sim \mathcal{U}(3, 6)$)

Fidelity definition

- **Quantity of interest:** Number of Responses/s
- HF: `ResponseSize` $100KB$ (average over 10 iterations)

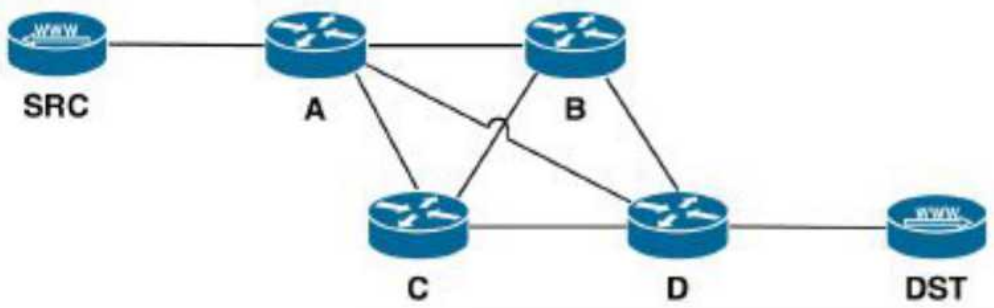
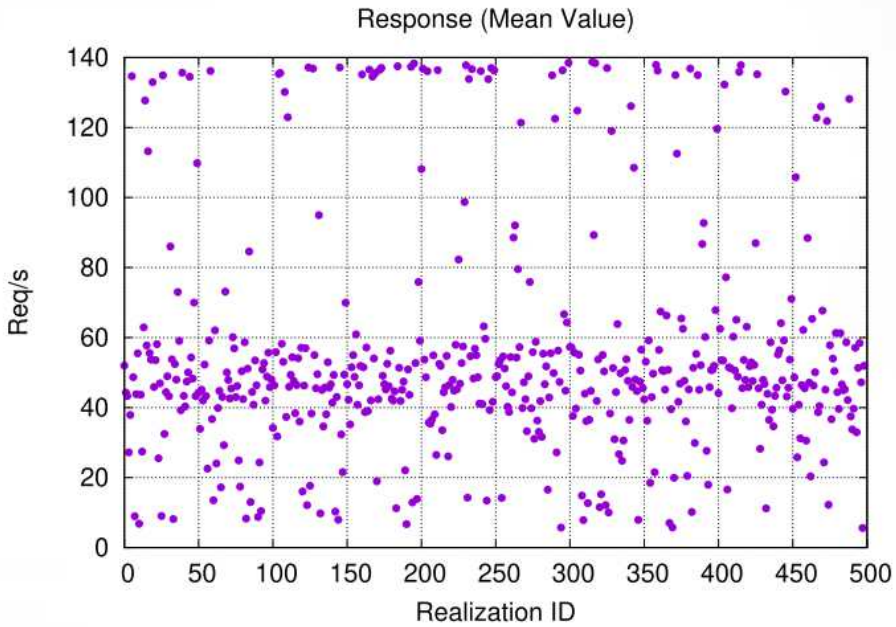
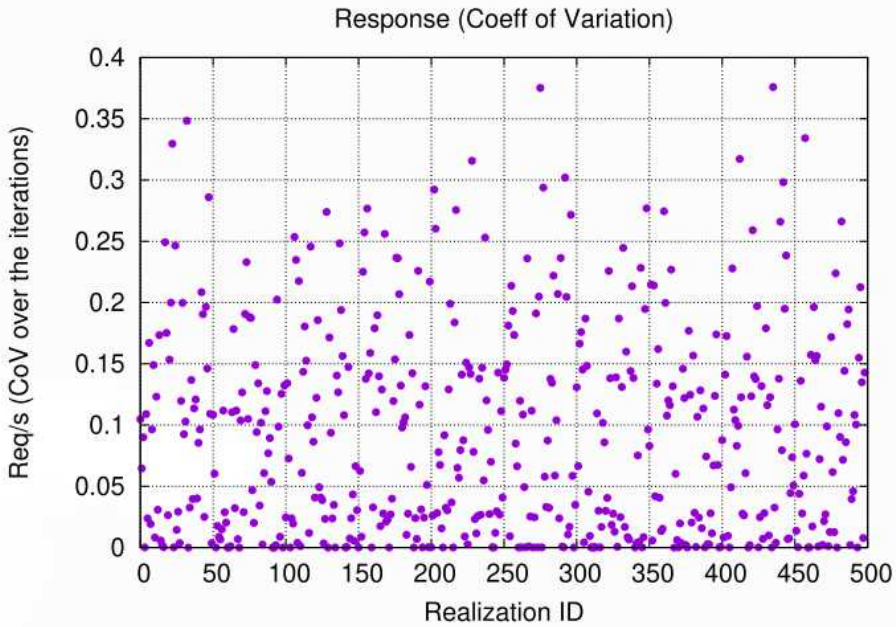


FIGURE: Network Configuration

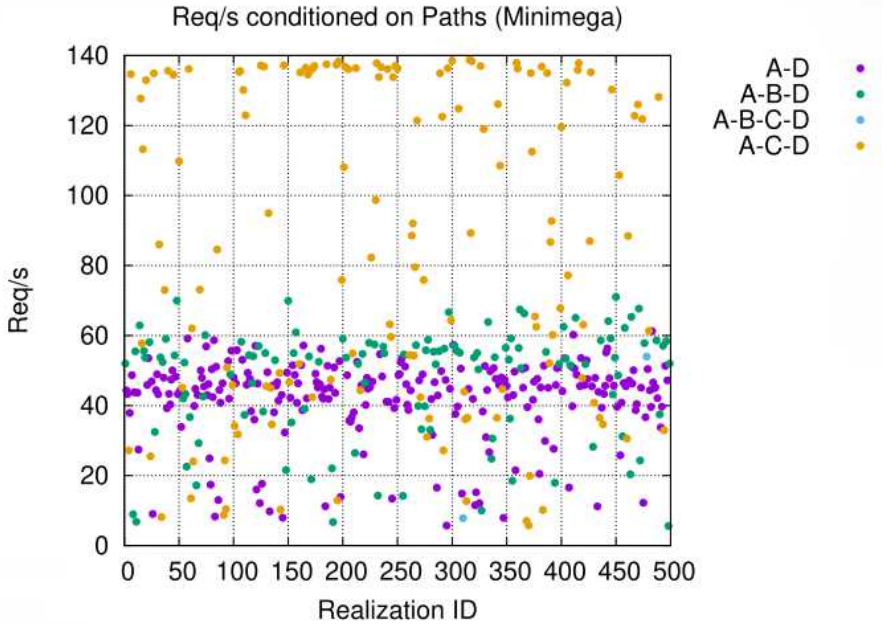
Requests/s – Network device: e1000

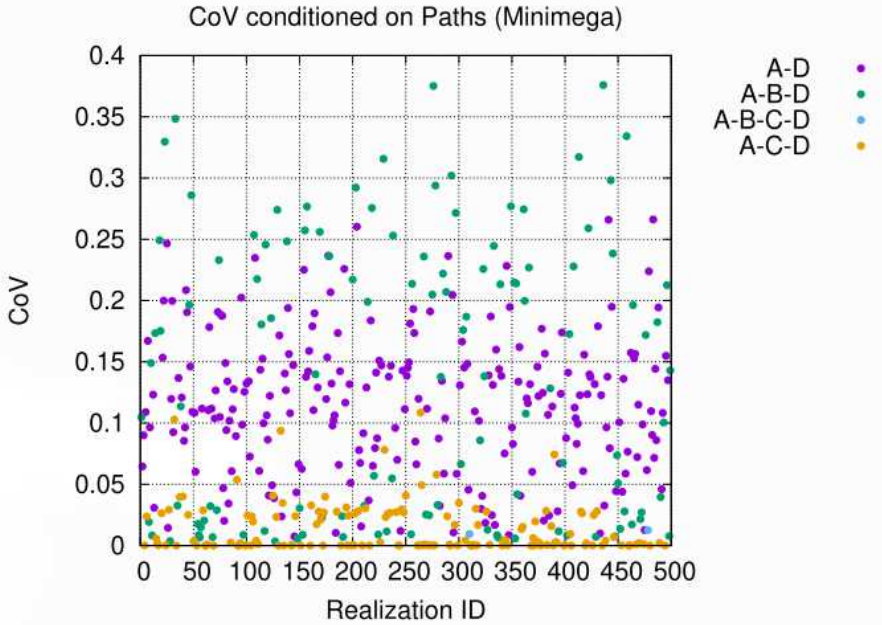




Can we study the response of the system conditioned on the paths?

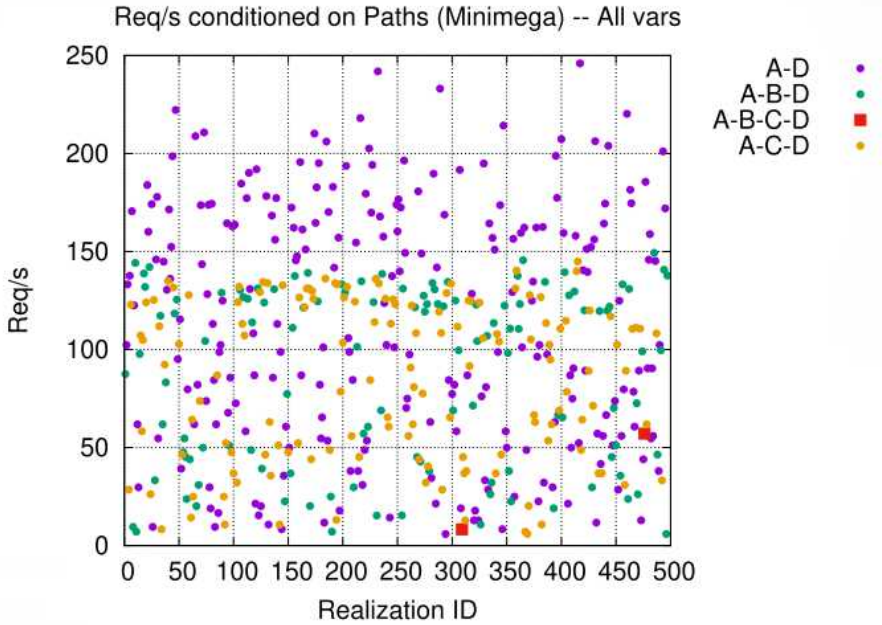
Requests/s – Network device: e1000

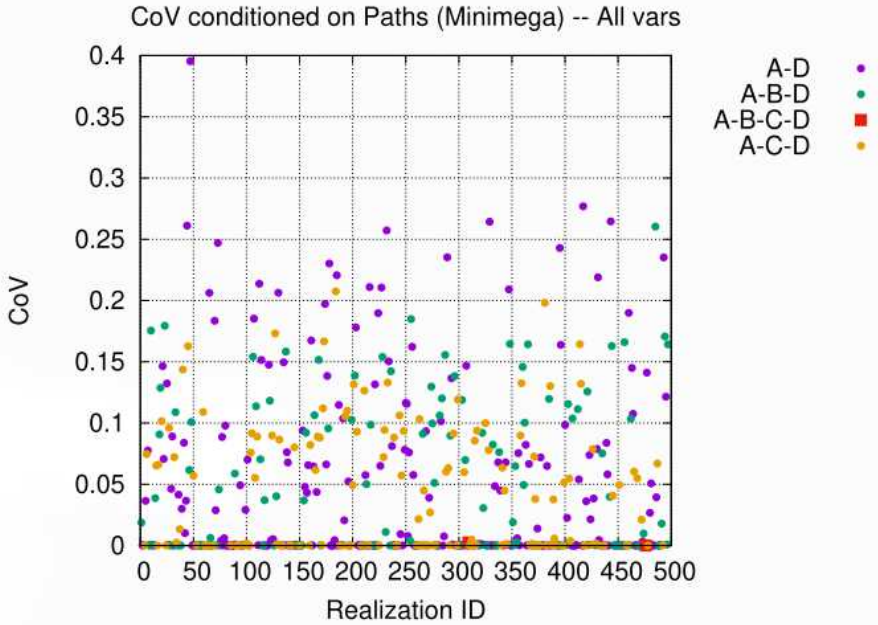




Why is the response over the paths A-B-D and A-C-D different? They should be consistent...

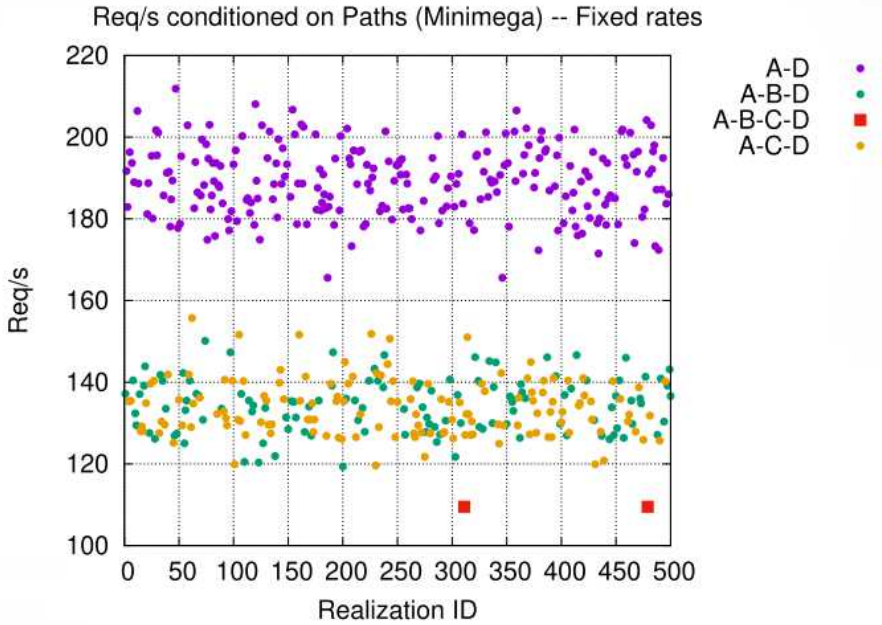
Requests/s – Network device: `e1000` → `virtio`

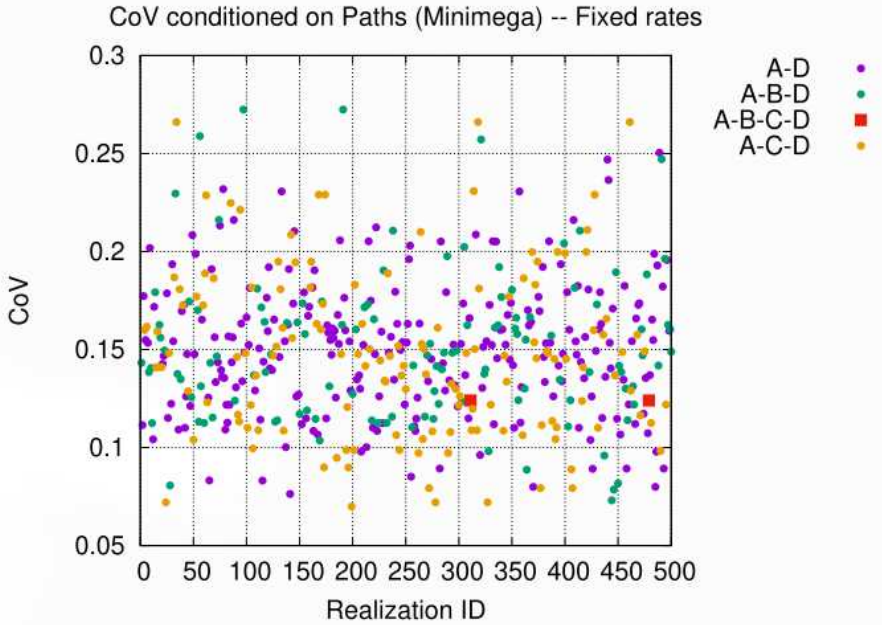




We looked at the network interface and replaced `e1000` with `virtio`

Requests/s – Network device: `e1000` → `virtio`





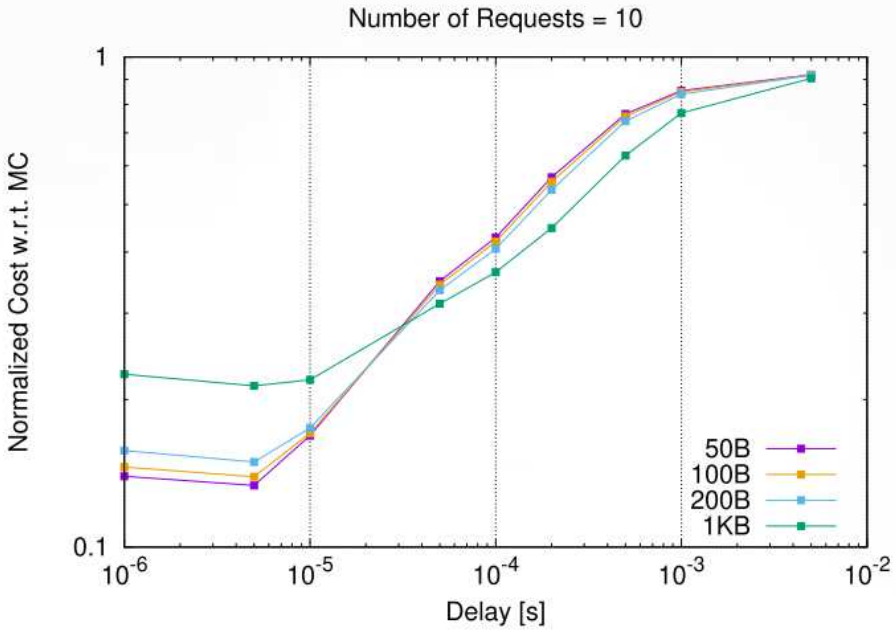
This further demonstrate the consistence of results over different paths

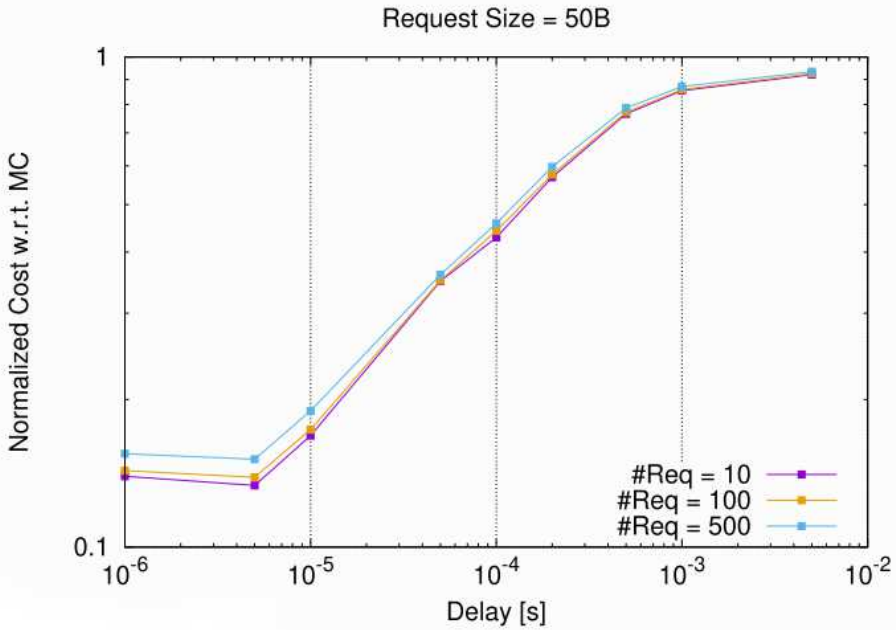
How difficult it is to select a low-fidelity model?

Can we use the flexibility in selecting the LF at our advantage?

How do we select the 'best' low-fidelity model?

- The LF model **does not need to be predictive** (*i.e.* the BIAS w.r.t. the HF can be very large), but we need it to be correlated and inexpensive to run
- Therefore, designing **a priori** a LF model might **not always be the best solution** (for MF)
- Very often a mismatch in parameterization exists between HF and LF $\rightarrow$ we can use the **'free' parameters as tuning parameters** to increase the correlation (given a finite set of HF data)

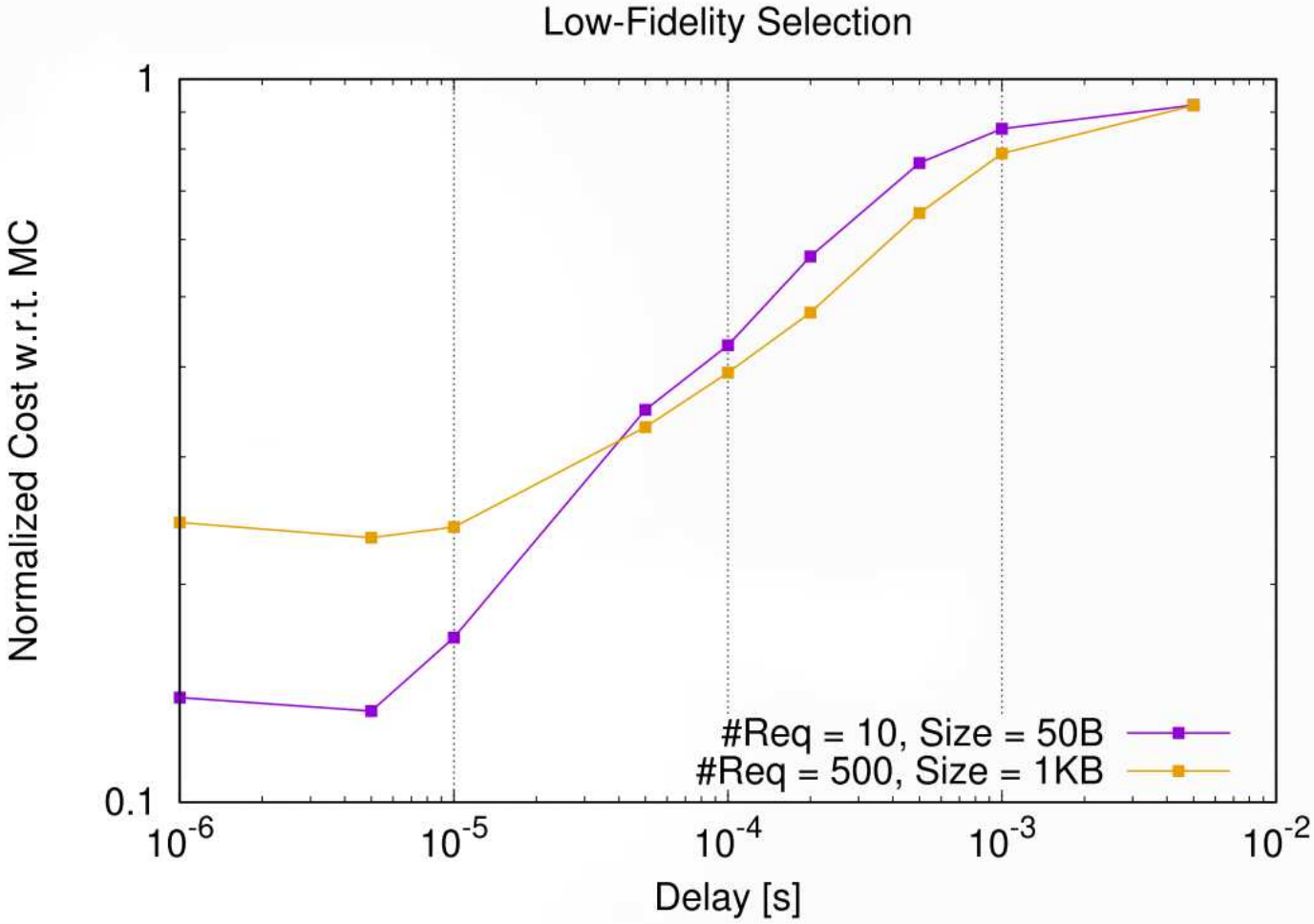




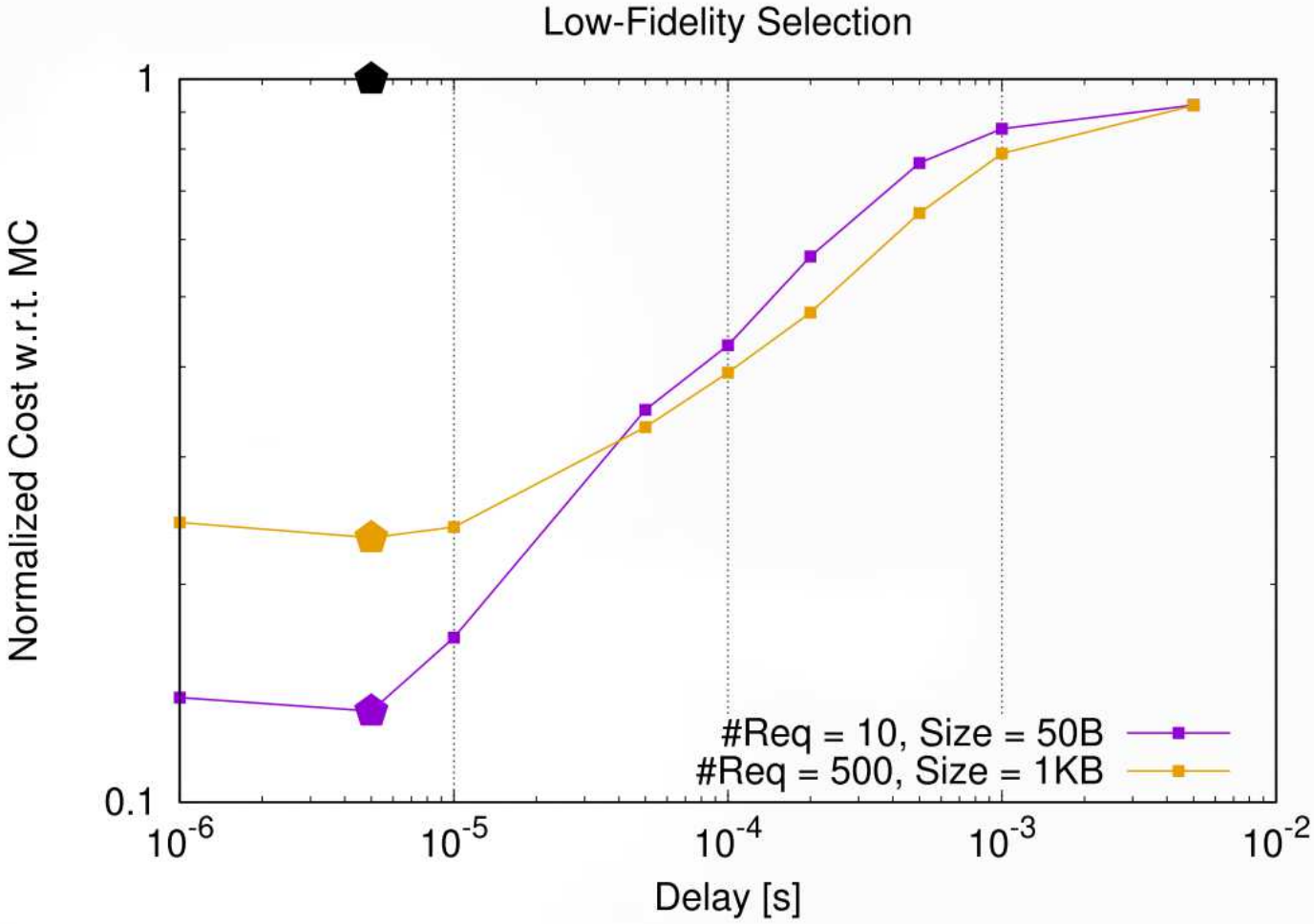
Notes:

- Several LF models can be obtained for different combinations of `Number of requests` and `Payload Size`
- Each LF combination has a different correlation (with `minimega`) and cost
- The cost of a MF estimator depends on the properties of the LF model

The LF can be optimized to obtain the maximum accuracy for a MF estimator beforehand, *i.e.* without requiring additional HF runs



The most predictive LF model on paper, might not be the most efficient from a MF perspective



The most predictive LF model on paper, might not be the most efficient from a MF perspective

How can we use this technology to support our customers?

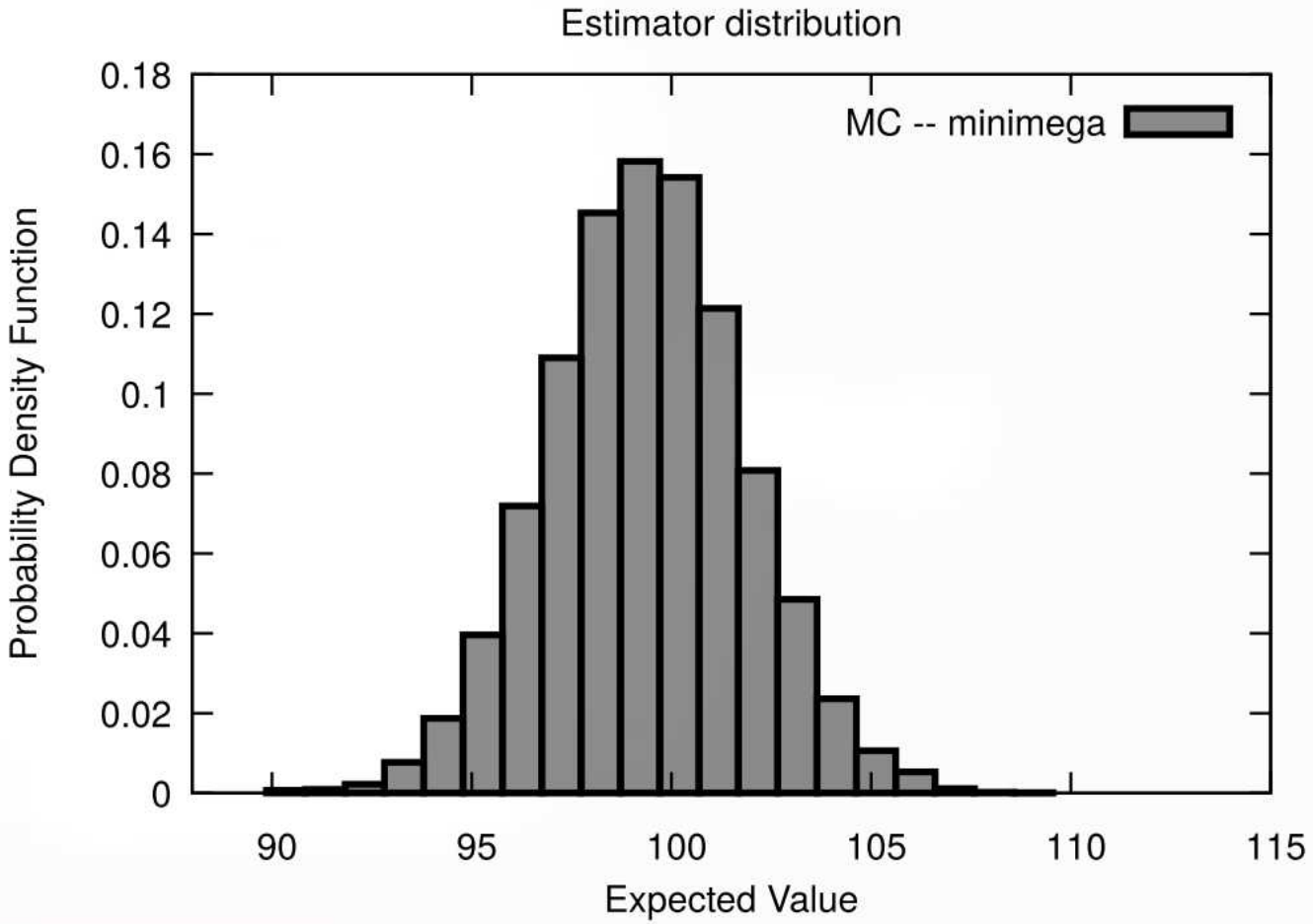
Customer X: I have a physical system for which I've collected data in the presence of uncertainty. Can you help me assessing the response's statistics?

Customer X: I have a physical system for which I've collected data in the presence of uncertainty. Can you help me assessing the response's statistics?

A generic UQer: Yes, give me the data and let's run a MC simulation

Customer X: I have a physical system for which I've collected data in the presence of uncertainty. Can you help me assessing the response's statistics?

A generic UQer: Yes, give me the data and let's run a MC simulation



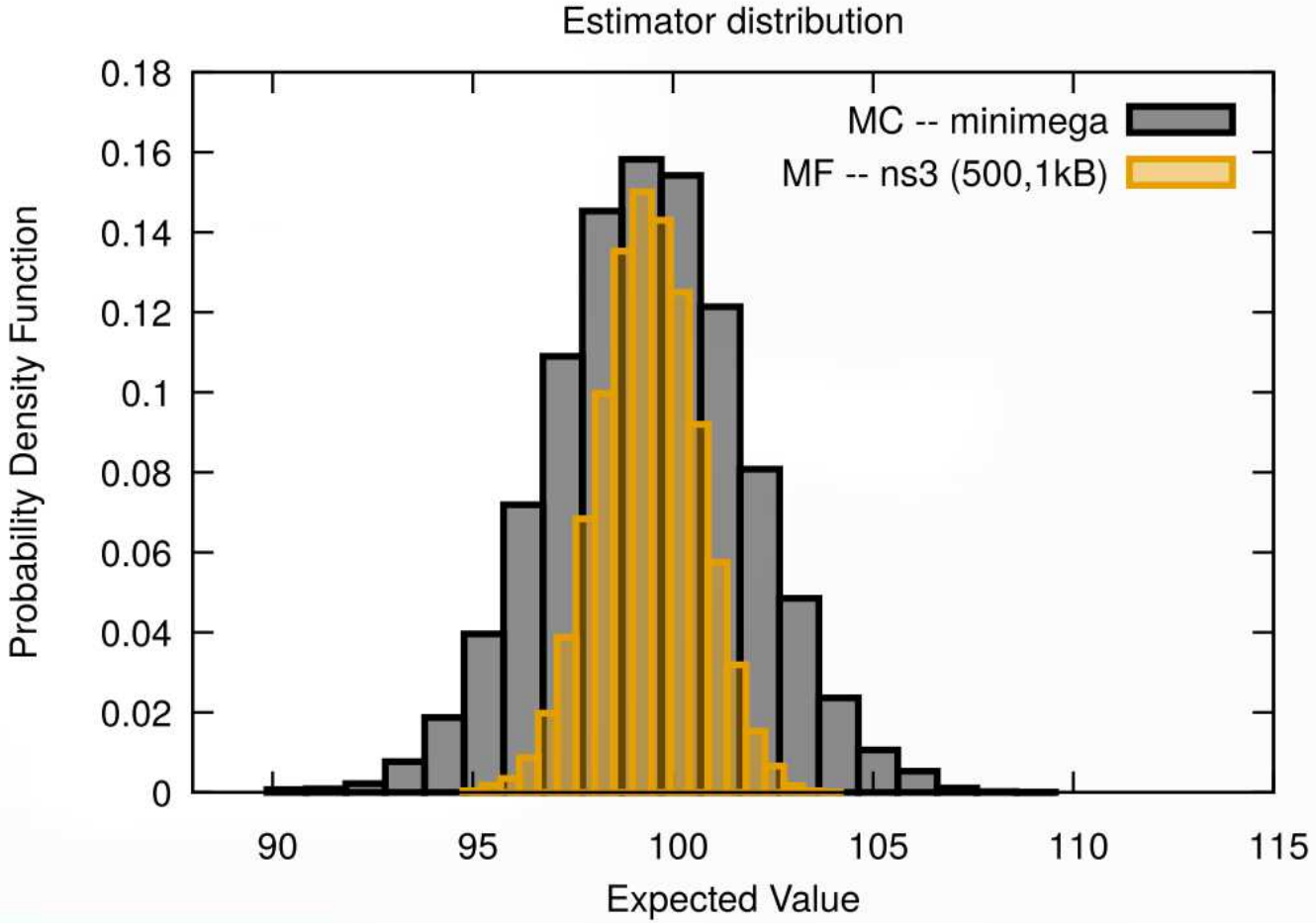
Customer X: I have a physical system for which I've collected data in the presence of uncertainty. Can you help me assessing the response's statistics?

Customer X: I have a physical system for which I've collected data in the presence of uncertainty. Can you help me assessing the response's statistics?

The MF-UQer: Yes, give me the data and let's run a MF simulation. Please also give me your **best** (*i.e.* most predictive) LF model

Customer X: I have a physical system for which I've collected data in the presence of uncertainty. Can you help me assessing the response's statistics?

The MF-UQer: Yes, give me the data and let's run a MF simulation. Please also give me your **best** (*i.e.* most predictive) LF model



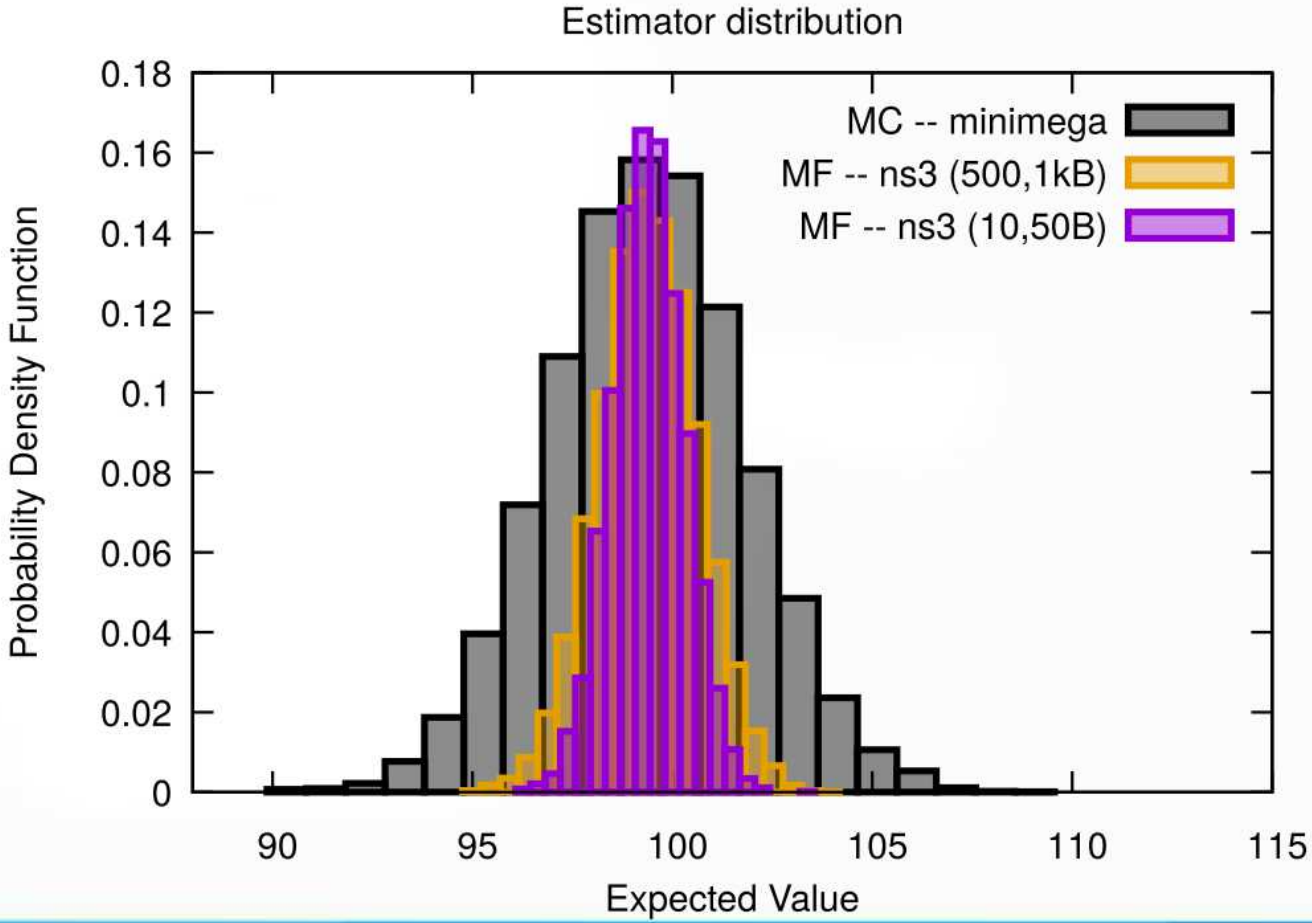
Customer X: I have a physical system for which I've collected data in the presence of uncertainty. Can you help me assessing the response's statistics?

Customer X: I have a physical system for which I've collected data in the presence of uncertainty. Can you help me assessing the response's statistics?

The SECURE-UQer: Yes, give me the data and let's run a MF simulation. Let me tune the LF model (no physical experiments)

Customer X: I have a physical system for which I've collected data in the presence of uncertainty. Can you help me assessing the response's statistics?

The SECURE-UQer: Yes, give me the data and let's run a MF simulation. Let me tune the LF model (no physical experiments)



Concluding Remarks

State-of-the-art

- Multifidelity Uncertainty Quantification proved to be effective for many different applications
- Encouraging results have been obtained for simplified network configurations and scenarios

Lessons learned from the Emulytics standpoint

- Configuration of the network devices (and potentially other parameters) has an impact on the system response $\rightarrow$ Validation
- Routing protocols used in the minimega and ns-3 models break ties in path costs differently

Future Directions

- Extension to additional statistics (Tails, risk measures, *etc.*)
- Multifidelity Sensitivity Analysis
- Exploration of data-driven approaches for LF modelling (ROMs, active directions, *etc.*)
- How to fully parameterize models? Scripts with 20+ arguments begins to get unruly. Topology generator $\rightarrow$ Dakota annotated graph $\rightarrow$ minimega/ns-3?
- How to pivot these experiments to more security-relevant quantities of interest? Study denial of service? Simulation and emulation may not be the best models for real-world DoS.

Approximate Control Variates

Optimal Control Variate

M low-fidelity models with known expected value

Let's consider M **low-fidelity models with known mean**. The **Optimal Control Variate (OCV)** is generated by adding **M unbiased terms to the MC estimator**

$$\hat{Q}^{\mathrm{CV}} = \hat{\mathbf{Q}} + \sum_{\mathbf{i=1}}^{\mathbf{M}} \alpha_{\mathbf{i}} \left(\hat{\mathbf{Q}}_{\mathbf{i}} - \mu_{\mathbf{i}} \right)$$

- $\hat{Q}_i$ **MC estimator** for the ith low-fidelity model
- μ_i **known expected value** for the ith low-fidelity model
- $\underline{\alpha} = [\alpha_i, \ldots, \alpha_M]^{\mathrm{T}}$ **set of weights** (to be determined)

Let's define

- The covariance matrix among all the low-fidelity models: $\mathbf{C} \in \mathbb{R}^{M \times M}$
- The vector of covariances between the high-fidelity Q and each low-fidelity Q_i: $\mathbf{c} \in \mathbb{R}^{M}$
- $\bar{\mathbf{c}} = \mathbf{c}/\mathbb{V}ar(Q)$, where ρ_i is the correlation coefficient (Q, Q_i)

The **variance of the OCV estimator** (optimal weights $\underline{\alpha}^{\star} = -\mathbf{C}^{-1}\mathbf{c}$)

$$\mathbb{V}ar\left(\hat{Q}^{\mathrm{CV}}\right) = \mathbb{V}\mathbf{ar}\left(\hat{\mathbf{Q}}\right)\left(\mathbf{1} - \mathbf{R}^{\mathbf{2}}_{\mathbf{OCV}}\right) = \mathbb{V}\mathbf{ar}\left(\hat{\mathbf{Q}}\right)\left(\mathbf{1} - \bar{\mathbf{c}}^{\mathrm{T}}\mathbf{C}^{-1}\bar{\mathbf{c}}\right), \quad 0 \leq R^2_{OCV} \leq 1.$$

NOTES:

1. For a single low-fidelity model: $R^2_{OCV-1} = \rho_1^2$
2. For all estimators in literature (MLMC, MFMC, *etc.*): $R^2 \leq \rho_1^2 \leq R^2_{OCV}$

LDRD

Laboratory Directed Research and Development

Optimization Thrust Overview

Team Members

Anya Castillo, Team Lead

Bryan Arguello

Jared Gearhart

William Hart

Emma Johnson (GATech)

Cindy Phillips

She'ifa Punla (RPI)

Presenter

William Hart

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.

Optimization Thrust Overview

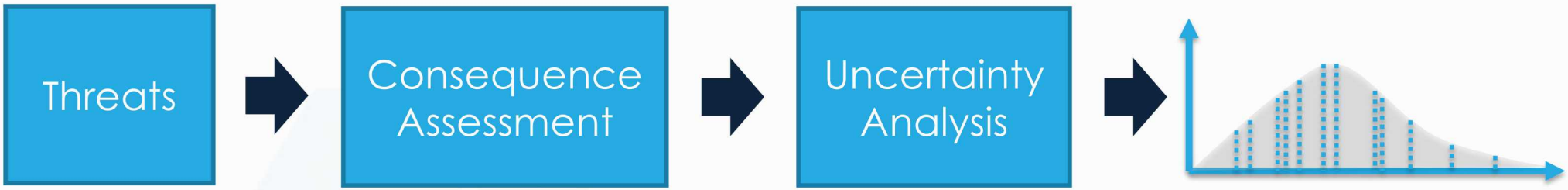
Problem: Decision-makers need to protect power grids against informed, adaptive, malicious adversaries attacking their cyber networks

Decision-makers need to:

- Account for likely adversarial behaviors/responses
- Plan response strategies
- Discover effective investment options

Challenge: There are **exponentially** many adversarial behaviors, response strategies, and investment options

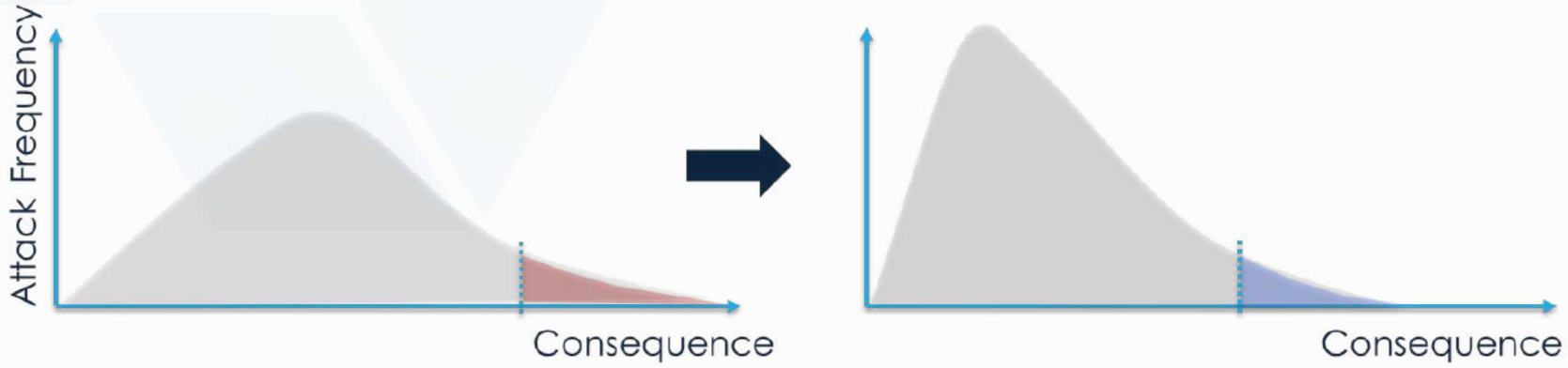
Cyber Risk Modeling in Grid Systems



Emulytics

Uncertainty Quantification

- **Optimization Goal**: Identify investment options that most effectively protect critical systems from cyber-physical threats



Motivating Concerns

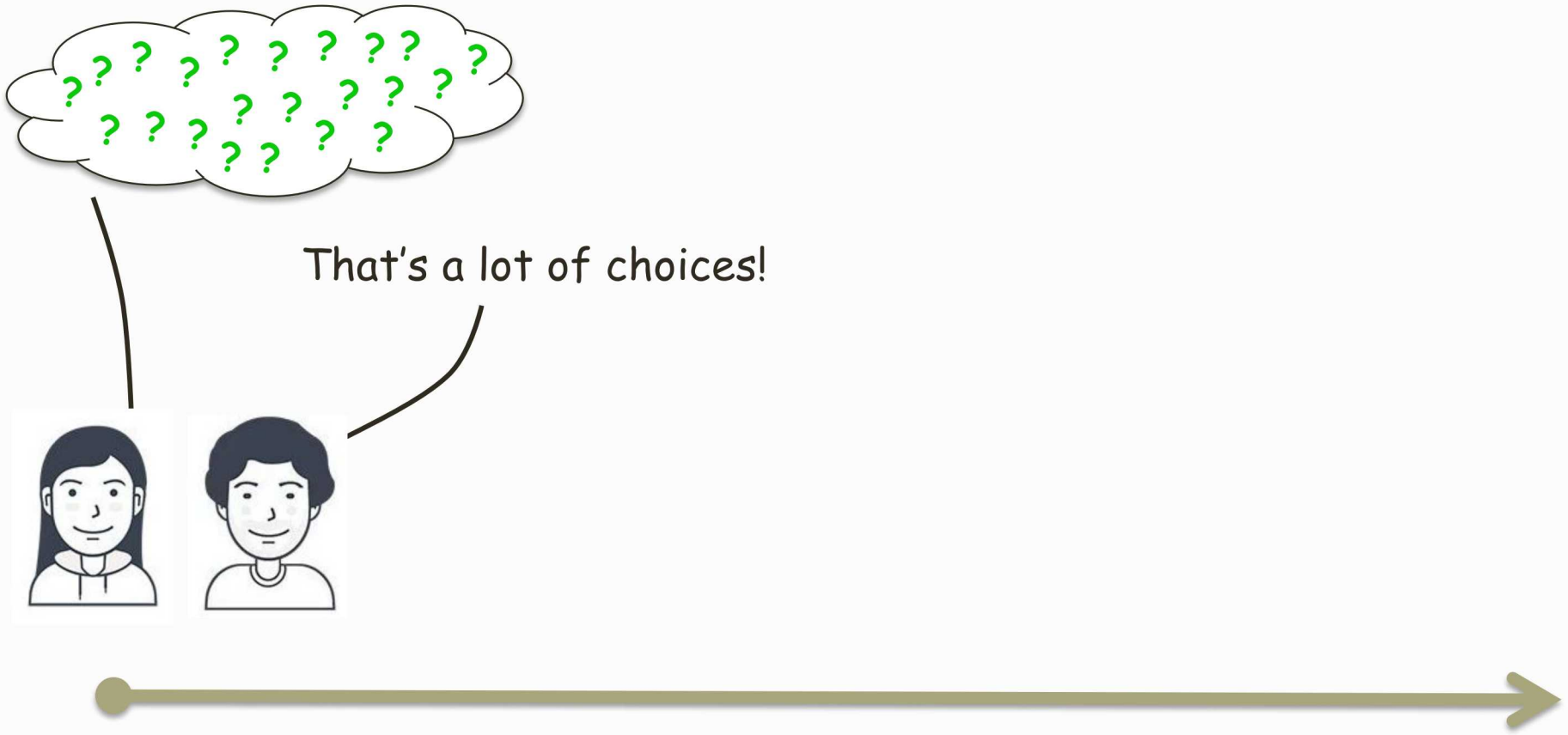
System Design is the focus of the optimization thrust

- How do we model system-level consequences?
- Where should we place cyber detectors in our networks?
- How can we partition our network to enhance security?

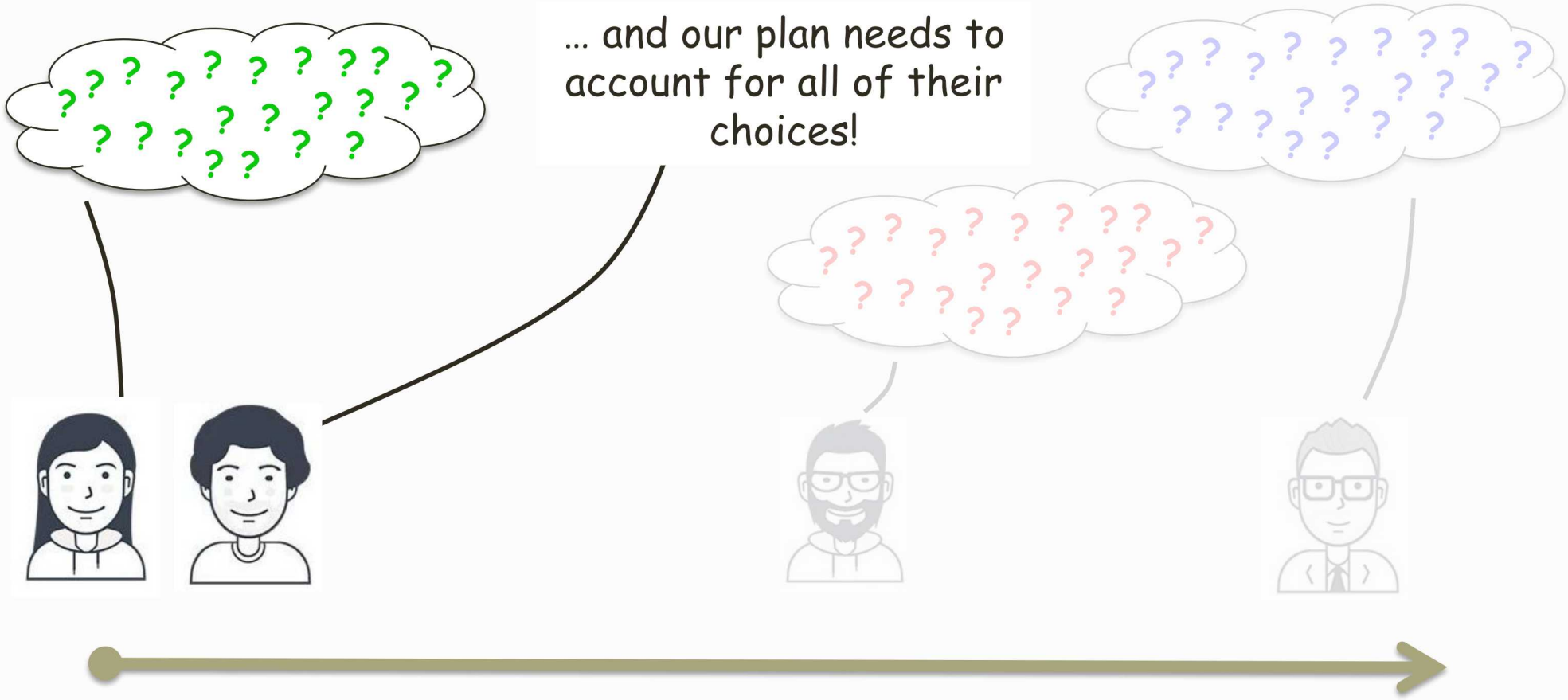
Key Challenges

- System models can have many parameters, but we will have limited data from Emulytics and UQ
- The optimization space accounts for threats and system options
  - The threat space is very large
  - Even small systems can have very large design spaces!

Interdiction – A Different Approach



Interdiction – A Different Approach



What's New?

Linear Programs

- Easily solved
- Widely used commercial and academic solvers

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

NOTE: These methods are not cyber or grid specific

Linear Bilevel Programs

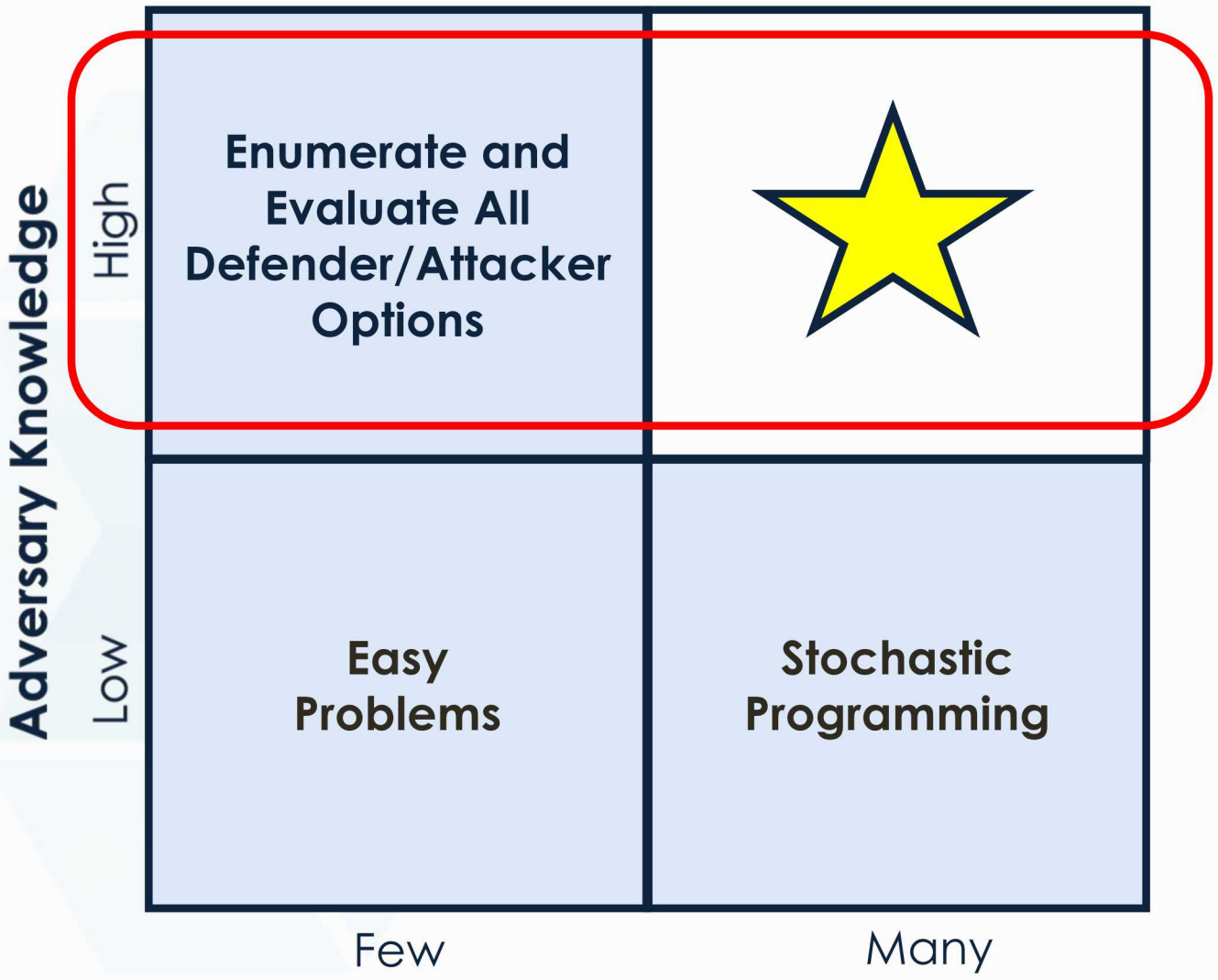
- Hard problems (NP-hard)
- No general-purpose commercial solvers for **discrete lower level decisions**

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

Upper Level Problem

Lower Level Problem

What's New?


Adversary Knowledge
High
Low
Enumerate and Evaluate All Defender/Attacker Options
Easy Problems
Stochastic Programming
Few
Many
# of Options to Defend and Attack

How Do We Solve These Problems?

Bilevel Solvers (Python)

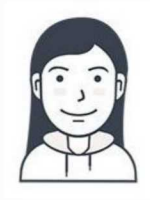
Opt Modeling (Python)

Opt Solvers (Python)

Opt Modeling (C++)

Bilevel Solvers (C/C++)

LP/IP/NLP Solvers (C/C++)



An Optimization Modeling Tool

- Built in Python
- Diverse modeling capabilities
  - Stochastic programs, disjunctive programs, etc
- Can express modular, hierarchical model structure
- Automatic model transformations

2016 R&D 100 WINNER

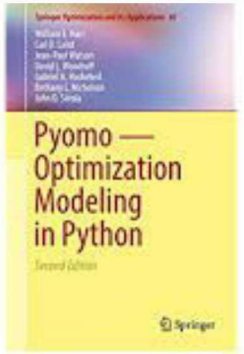
INFORMS Computing Society Prize - 2019

> 130,000 software downloads in 2019

> 90,000 chapter downloads of the Pyomo book





Technical Focus Areas

	FY19	FY20	FY21
Defender-Attacker Game Models	Bi-Level (Attacker-Defender)	Multi-Level (Defender-Attacker-Defender)	Generalization of Formulation Approach
Knowledge and Data Uncertainty	Models of Cyber Components	Stochastic Interdiction	Distributionally Robust
Tailored Optimization Algorithms	Algorithms for Nonconvex Subproblems	Algorithms for Discrete Subproblems	Global Methods for Nonconvex, Discrete Subproblems

Cyber Application in FY19: Worst-case scenario analysis
Cyber Application in FY20: Optimal Placement of Intrusion Detection
Cyber Application in FY21: Attack Graphs With Defender Intervention

- **Modeling (+), Algorithms (-)**
- **General Algorithms (+), Tailored Algorithms (-)**
- **Software (+)**

Major FY19 Developments

1. Assessing the state-of-the-art
2. Bilevel software and solvers
3. New cyber-grid models

1. Assessing the State-of-the-Art

- The optimization team has performed a review on the literature to understand the current state-of-the-art

- The team attended the International Workshop on Bilevel Programming (June, 2018)

The team decided to not focus on drafting an article for this review

  - It is not clear how general existing methods are

  - We are still learning what we need for cyber grid applications

Partial Bilevel-Optimization Survey

- Mixed or pure integer in both upper and lower, no stochasticity
  - [DeNegre and Ralphs 2008], [Domínguez and Pistikopoulos 2010], [Fischetti et al. 2016, 2017, 2018], [Kleniati and Adjiman 2015], [Lozano and Smith 2017], [Mitsos 2010], [Tahernejad et al. 2016], [Tang et al. 2016], [Wang and Xu 2017], [Wiesemann et al. 2013], [Xu and Wang 2014], [Yue et al. 2019], [Zeng 2015], [Zheng et al. 2018]
- No integer variables, no stochasticity
  - [Zheng et al. 2018], [Dempe et al. 2018]
- [Zhao and Zeng 2012] mixed-integer upper and lower with a notion of uncertainty. No probabilities, consider worst case. Tri-level
- Surveys
  - [Dempe 2005, 2018], [Liu et al. 2018]

Motivating Concerns

- Available software (Ralphs et al. and Fischetti et al.)
- Lower level integer decisions (Yue et al.)

Existing Branch and Cut Bilevel Solvers

- **MibS**
  - Ralphs et al. (Lehigh)
  - COIN-OR bilevel programming branch-and-cut solver
  - MibS is open source, we can look under the hood and add our own ideas
  - We have tested MibS on existing sample problems

- **"Fischetti Solver"**
  - Fischetti et al. (U. Padua)
  - Uses CPLEX branch-and-cut algorithm and built-in callbacks to make the branch-and-cut tailored to solve bilevel programming models
  - Leverages commercial solvers, which are likely to be robust
  - We have tested the solver on existing sample problems

- **Future Work**
  - Apply both solvers to our past models for validation
  - Determine which of our new models can be solved with these solvers

Zeng Solver

"Zeng Solver"

- A projection-based reformulation and decomposition algorithm
- Allows for the solution of bilevel programs with integer variables in the lower-level problem.
- Uses column-and-constraint-generation method to avoid enumerating all possible integer solutions.

Implementation in Pyomo

- Pyomo implementation leverages unique Pyomo capabilities
- This implementation runs successfully and correctly solves simple problems that were demonstrated in the paper.

Next steps

- Scalability studies
- Investigate the effect of adding uncertain data in the upper level constraints and solving robustly

2. Bilevel Software and Solvers

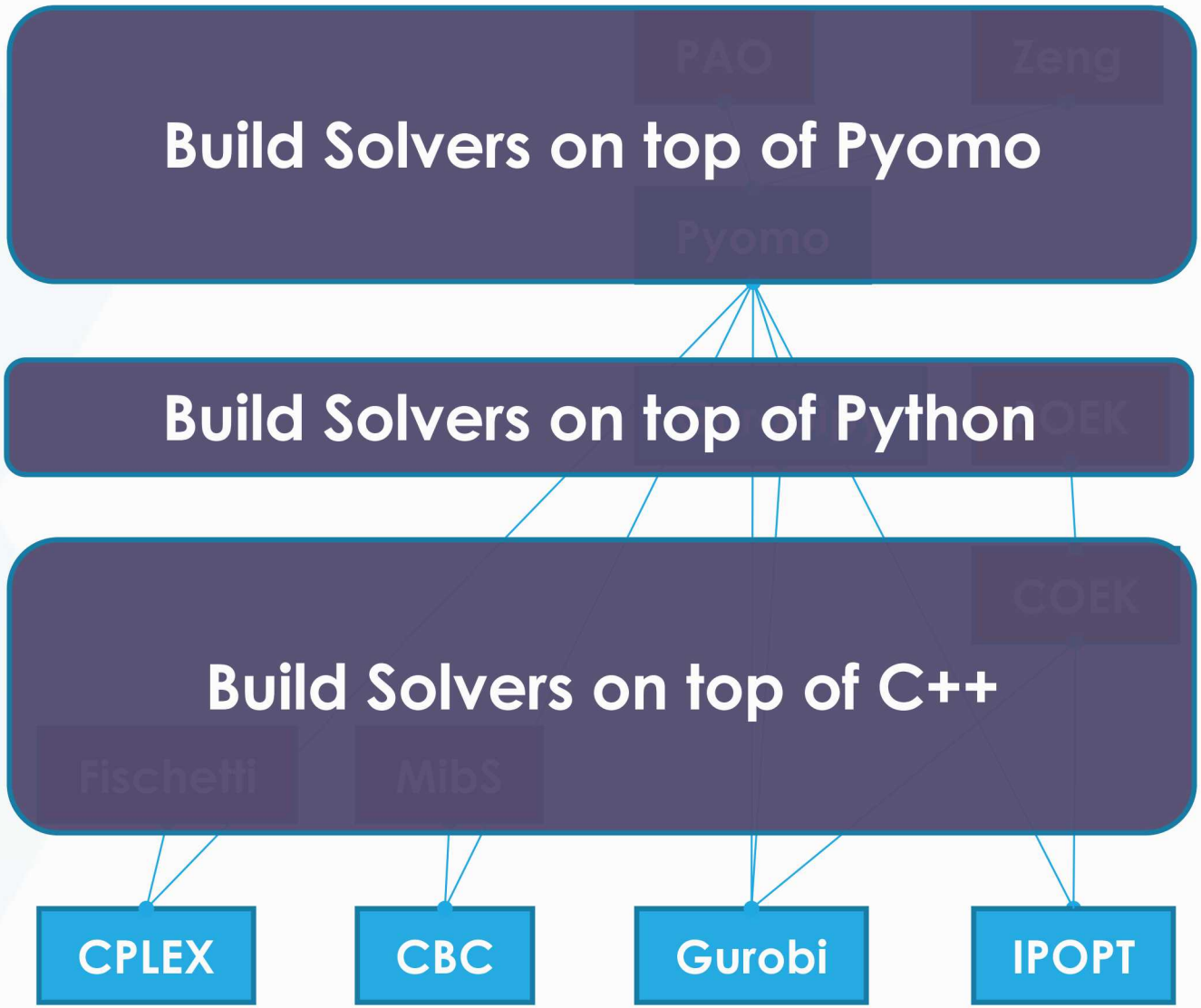
Challenge: how are we going to solve cyber-grid optimization applications in SECURE

- No commercial solvers exist for our applications
- Few academic solvers, with nontrivial limitations

Observation: This is an emerging area with a lot of interest

- When we deprecated *pyomo.bilevel*, people complained!

What is the "best" way to build solvers?



FY19 Software Activities

Using Pyomo

- Developed the PAO package from pyomo.bilevel
  - Reworked and generalized dualization logic
- Implemented bilevel solver of Zeng et al.

Using Python

- Considered implementing bilevel solver of Zeng et al.
- Developed POEK, which is 4-6x faster than Pyomo in problem setup
- Demonstrated that POEK supports fast resolves to Gurobi and IPOPT

Using C++

- Developed COEK, which supports solver-agnostic direct interfaces
- Interfaced COEK with Python (POEK)

Comparison of Approaches

Key Features	Using Pyomo		Fischetti Python		Fischetti C++		MibS C++
	PAO	Zeng	Gurobipy	POEK	COEK	CPLEX	COIN-OR
Fast Solvers	N		-	-	Y	Y	Y
Solver Agnostic	Y		N	Y	Y	N	Y
Expression Repn	Y		Y	Y	Y	Y	N
Matrix Repn	N		N*	N*	N*	Y	Y
Robust Xforms	Y		N	?	?	N	N

Optimization Software Strategy

Observations:

- Software has not dominating our work
  - This was a concern at the last review
- But, we need to make optimization software a larger focus in SECURE

Challenges With Existing Techniques:

- Software control is limited
- Minimal debugging information
- Licensing issues
- Fixed MIP solvers hard-coded (which has performance implications)
- Modeling limitations

FY20 Focus: Develop solver implementations to demonstrate that we can effectively deploy scalable solvers

3. New Cyber-Grid Models

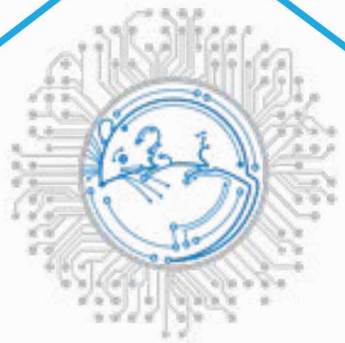
Challenge: Robust predictions with limited Emulytics/UQ predictions

Challenge: Capturing meaningful abstractions for Emulytics analyses

Models Parameterized with Emulytics Data

Optimization

Models that Help Define Emulytics Experiments

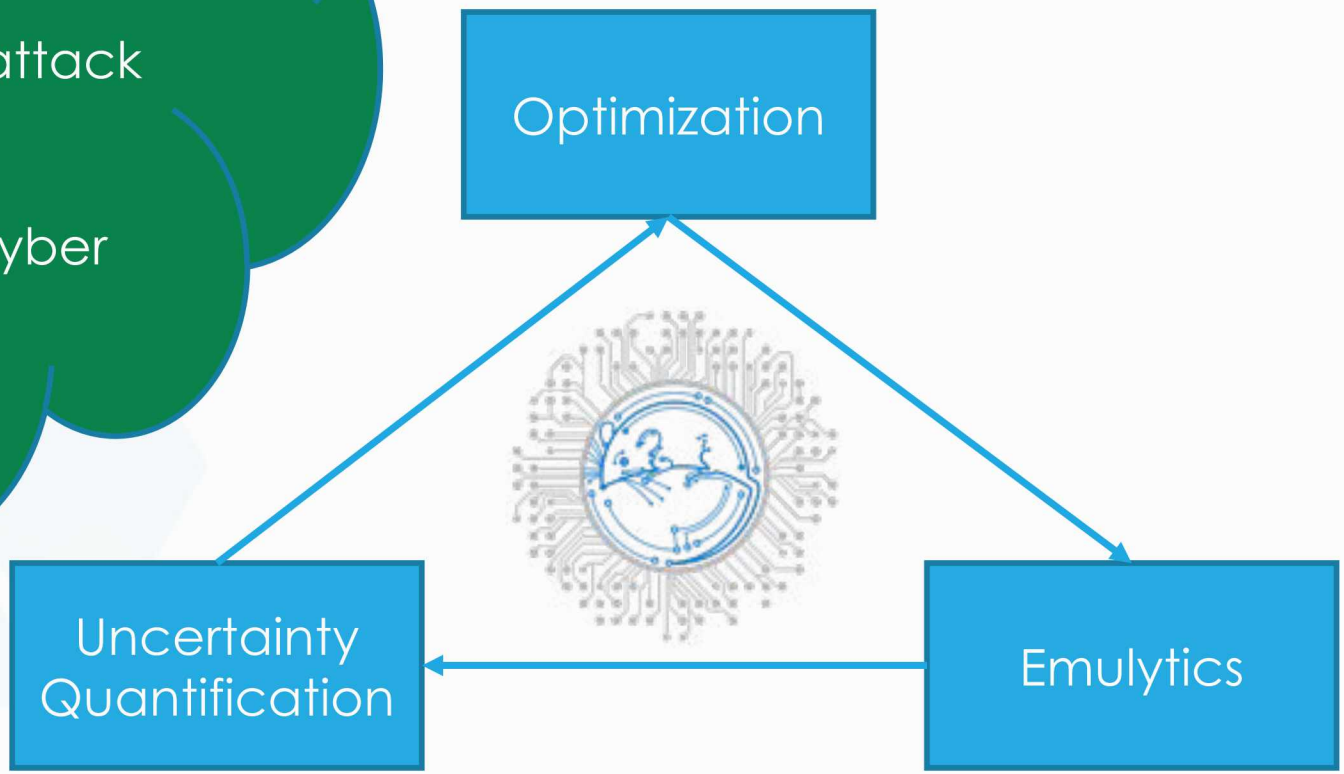


Uncertainty Quantification

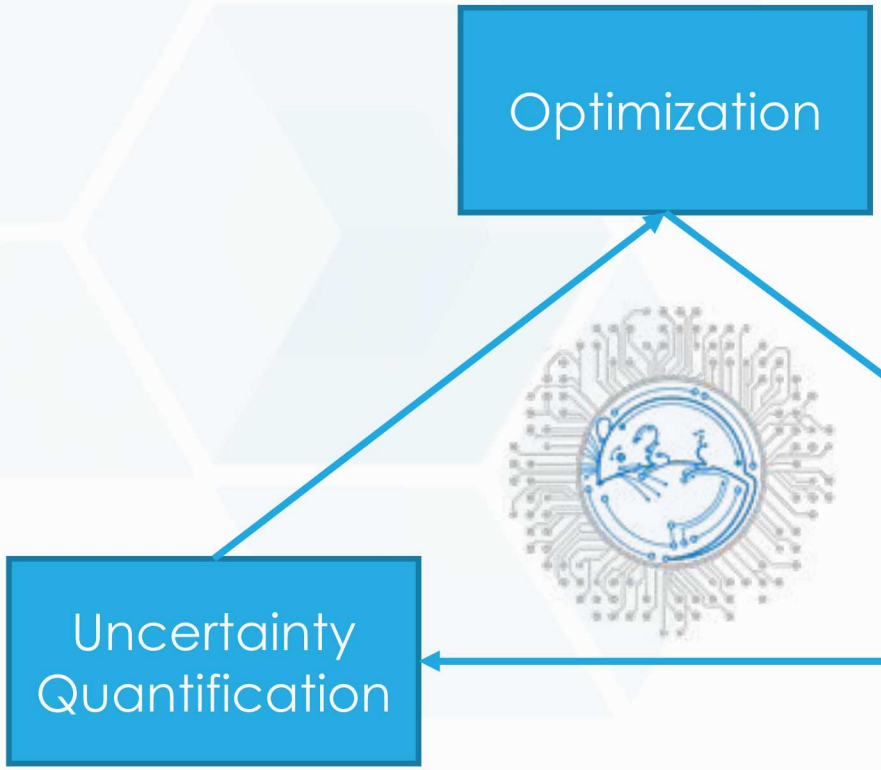
Emulytics

3. New Cyber-Grid Models

- Reachability with a simple topology-based attack model
- Modeling with attack graphs
- Placement of cyber sensors

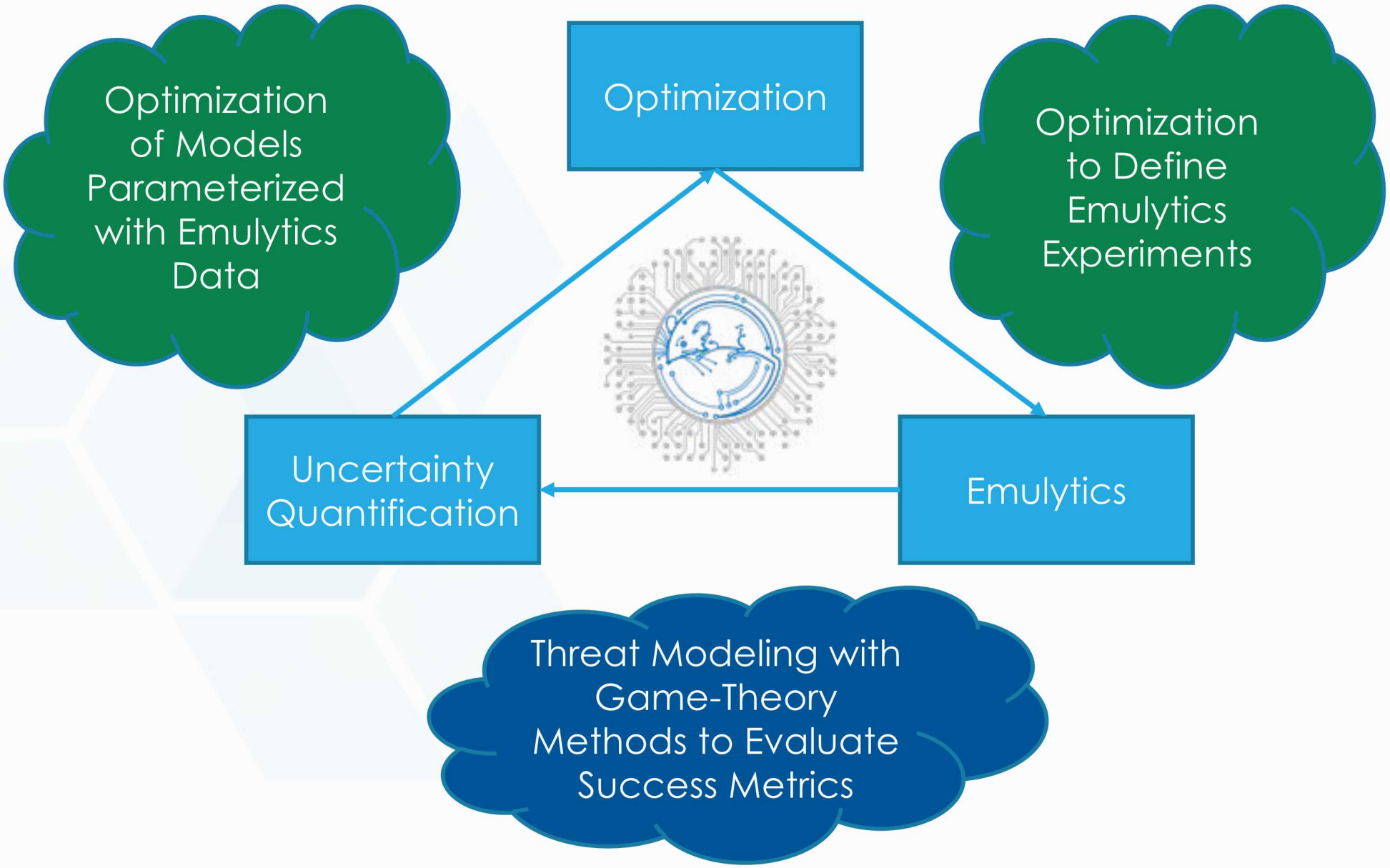


3. New Cyber-Grid Models



- Network segmentation to mitigate large disruptions
- *Assign physical devices to RTUs mitigate large disruptions*

Game Theory / Optimization



GPLADD / Bilevel Programming

Game Theory

- Can model multi-stage games
- Solvers often provide heuristic solutions
- Can provide exact solutions in special cases

Bilevel Programming

- Usually limited to 2- or 3-stage games
- Solvers usually provide exact solutions to general classes of problems
  - But not guaranteed to find solutions quickly
- Can provide bounds on optimality

FY20 Focus Areas

- Develop and refine cyber-grid models, with a focus on integration with UQ and Emulytics teams
  - Demonstration Problem focused on Intrusion Detection System Design
  - Demonstration problem focused on partitioning and device mapping

- Software development to enable the solution of these problems
  - Resolve issues using MibS (with Lehigh)
  - Re-implement Fiscetti et al approach using GUROBI
  - Address performance bottlenecks (in Pyomo or PAO)

- New algorithmic development
  - Robust formulations (RPI)
  - Pessimistic formulations (GATech)

Academic Collaborators
Dey, GATech
Mitchell, RPI
Ralphs, Lehigh
Zeng, U Pittsburg

Specific Accomplishments

- 6 Posters/Presentations
  - IMA COIN-OR Workshop, INFORMS Computing Society Conference, GraphX, IWOBIP Workshop, INFORMS Annual Meeting, Resiliency Week
- 2 Publications
  - One lead by optimization team
- Preliminary model implementations
  - Stochastic Worst Case Attacker
  - Stochastic Intrusion Detection Placement
  - Network Segmentation
- Preliminary solver implementations
  - PAO Dualizations
  - Zeng bilevel solver
- Copyright Assertions (in process)
  - PAO, COEK, POEK

LDRD

Laboratory Directed Research and Development

Cyber Physical Optimization Modeling

Team Members

Anya Castillo, Team Lead

Bryan Arguello

Jared Gearhart

William Hart

Emma Johnson (GATech)

Cindy Phillips

She'ifa Punla (RPI)

Presenter

Bryan Arguello

U.S. DEPARTMENT OF ENERGY

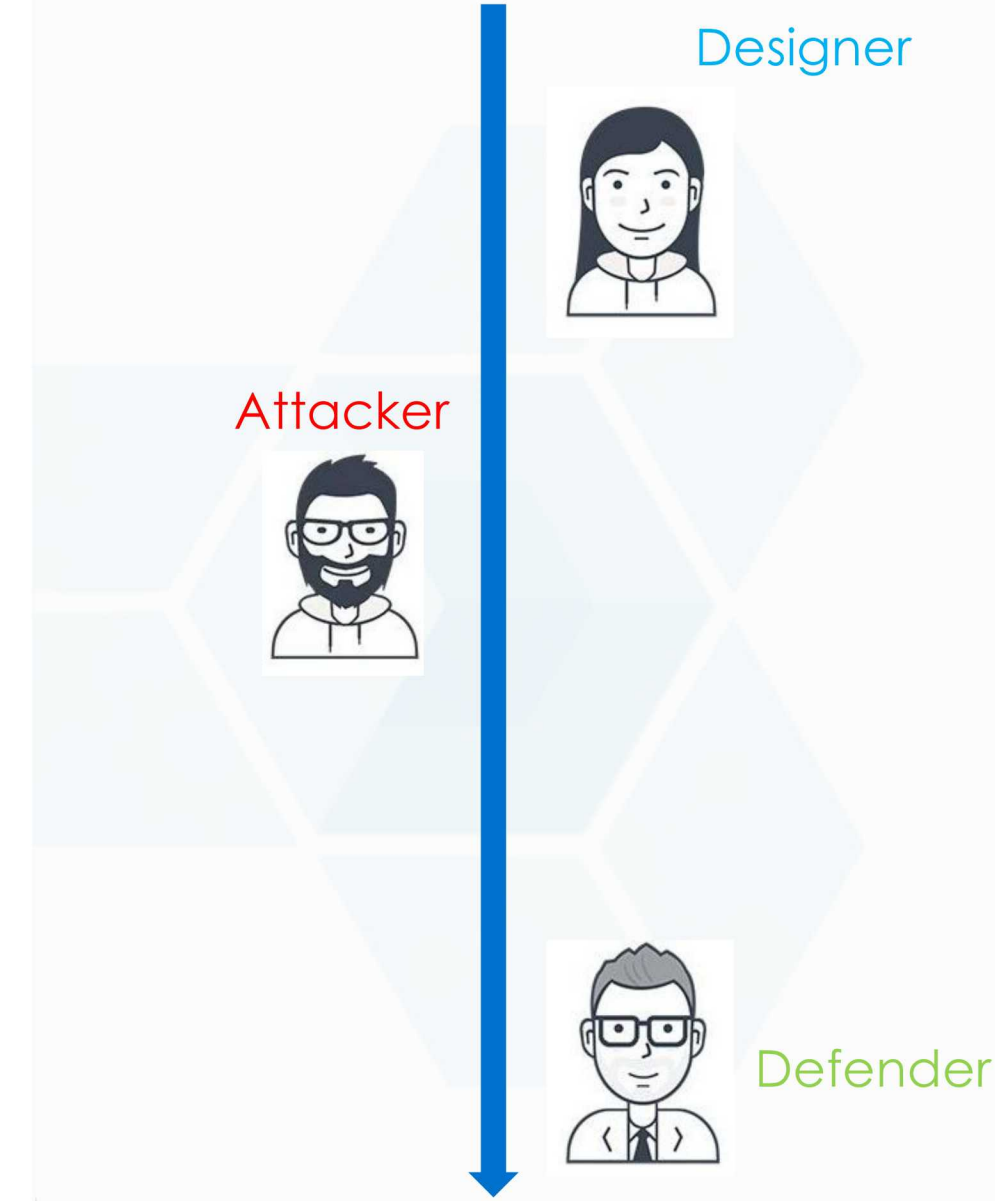
National Nuclear Security Administration

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.

Optimization Modeling Outline

- Trilevel Programming High-Level Overview
- Preliminary Cyber Physical Security Models
  - Worst Case Attacker
  - Stochastic Worst Case Attacker
  - Stochastic Intrusion Detection Placement
  - Network Segmentation
- Future Optimization Problems

Trilevel Programming



Meet the players!

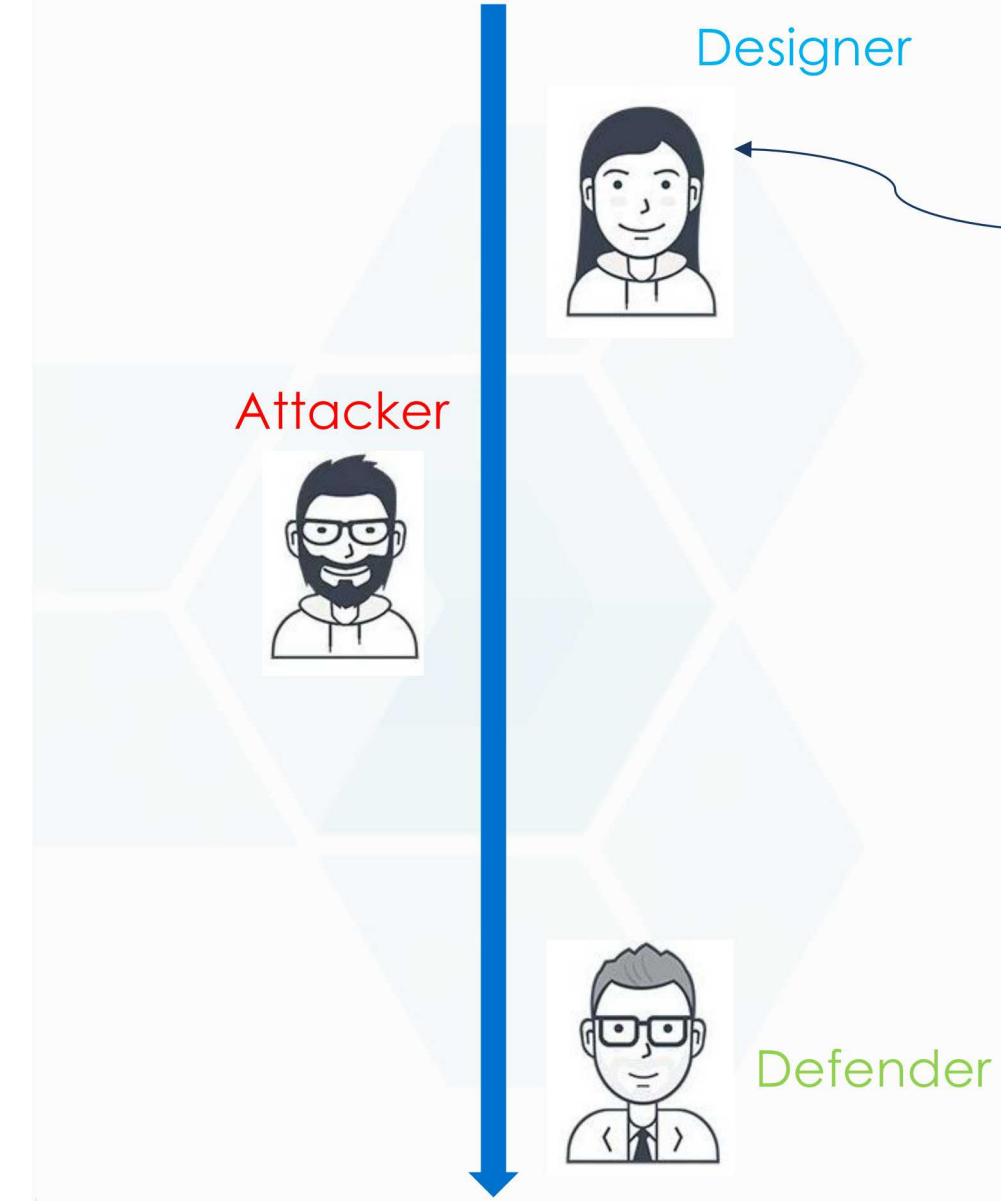
Each player gets to make exactly one set of decisions.

Designer goes first and knows what attacker will do

Attacker accepts designers decision and goes next knowing what defender will do

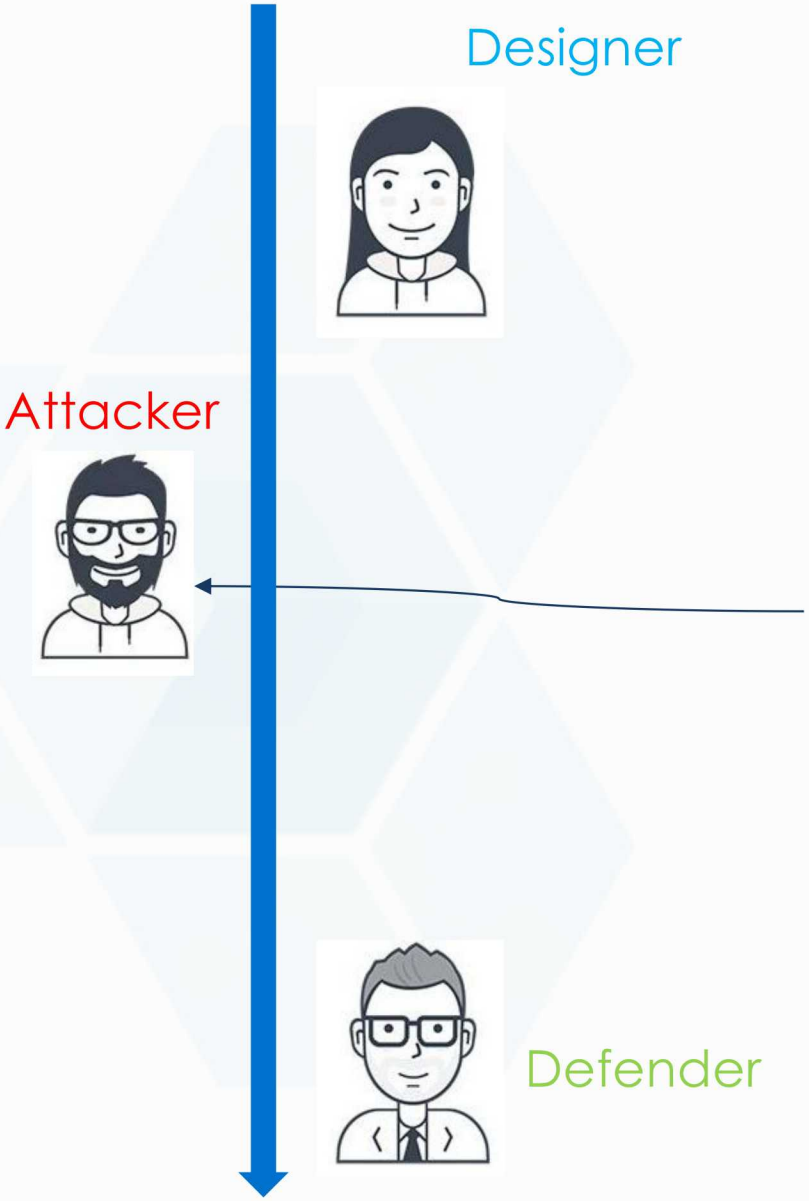
Defender goes last and accepts attacker decisions

Trilevel Programming



- Minimizes designer objective subject to design choice constraints
- Objective and constraints may include both designer and attacker decisions
- The designer knows what optimal choice attacker will make given design choices

Trilevel Programming

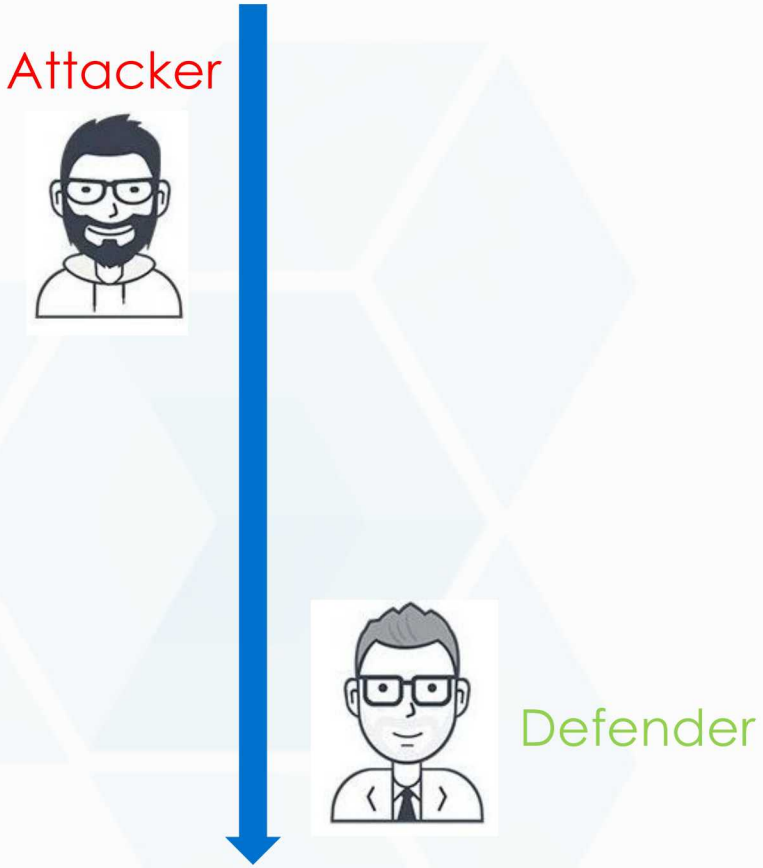


- The attacker accepts designer choices and must work with them. These designer decisions set the attacker's choices
- Maximizes attacker objective subject to attack choice constraints
- Objective and constraints may include both attacker and defender decisions
- The attacker knows what optimal choice defender will make given attack choices

Trilevel Programming

- The Defender accepts attacker choices and must work with them. These attack decisions set the defender's choices.
- Optimizes an objective that depends on only defender decisions
- Constraints includes defender decisions only

Bilevel Programming



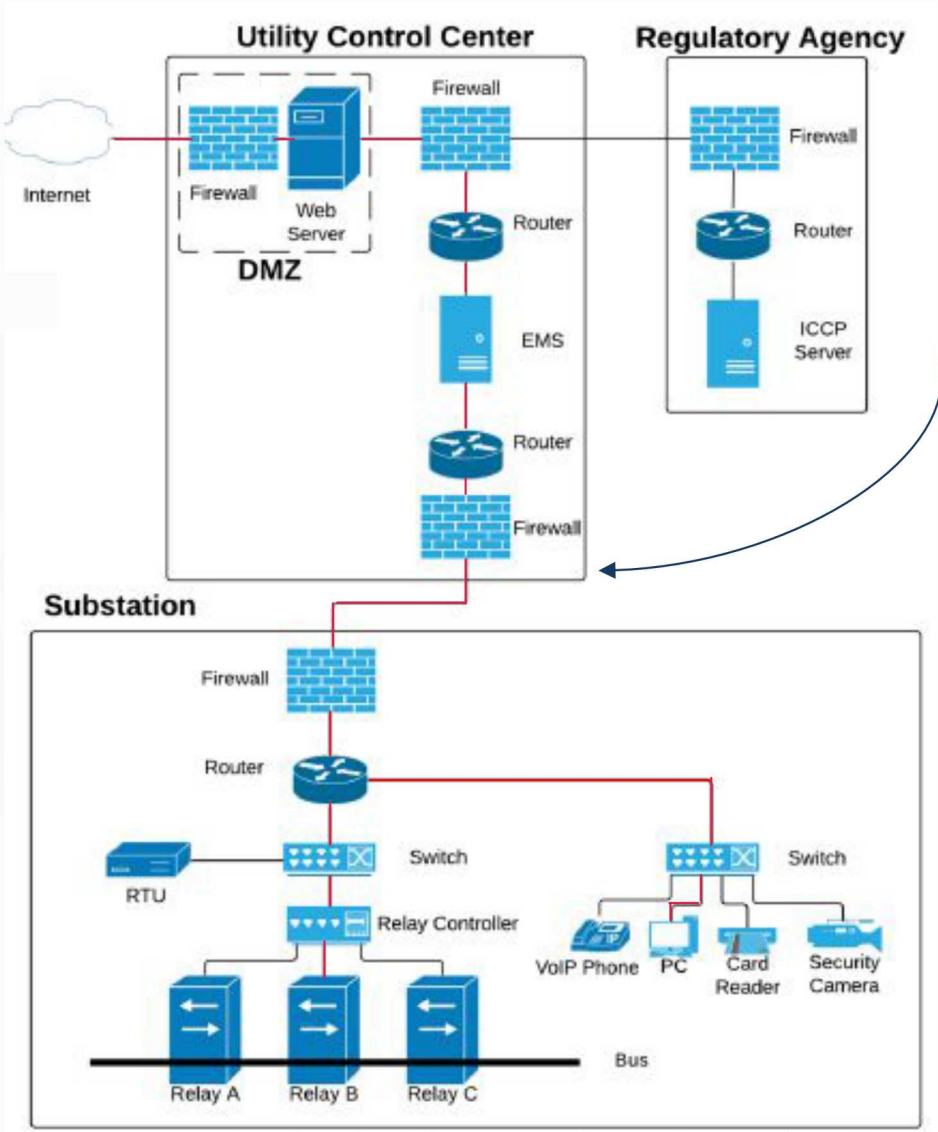
Same game, one less player!

Feedback Request

Now I will present a series of four bilevel/trilevel models for addressing cyber physical security questions.

I welcome any feedback you can give on these models!

Cyber Physical Attack Sequence Modeling



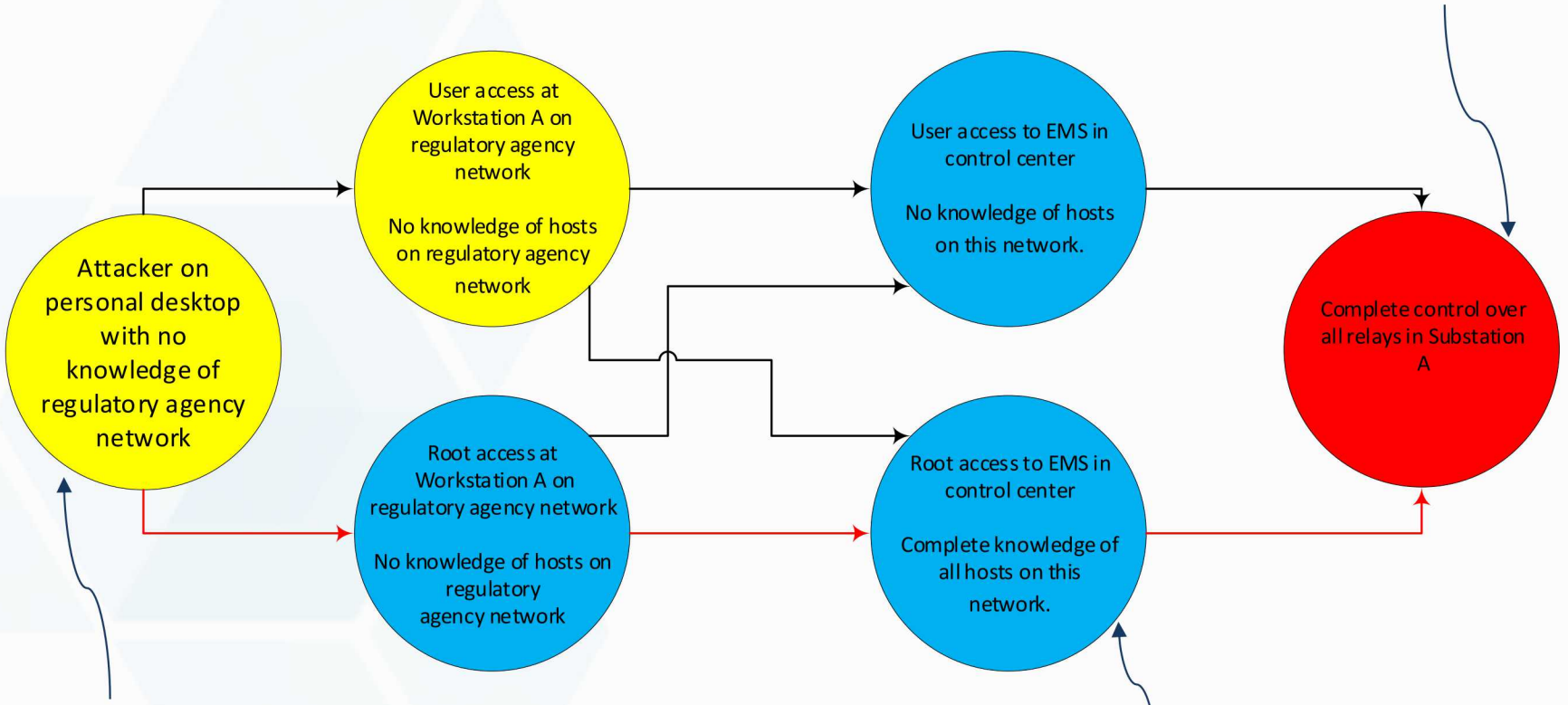
- Elements of cyber attack sequence
  - Sequence of hosts
  - Attacker access at hosts
  - Attacker actions at hosts
  - Network knowledge
  - Success probabilities
- Consider multiple attack sequences with some overlapping effort
- First question: which attack sequences are most damaging to the grid?

Oleg Sheyner and Jeannette Wing. Tools for generating and analyzing attack graphs. In Formal Methods for Components and Objects, volume 3188, pages 344–371. Springer, 2003.

Attack Graph

A simple example with
6 attack sequences…

Terminal nodes inflict damage on grid if reached



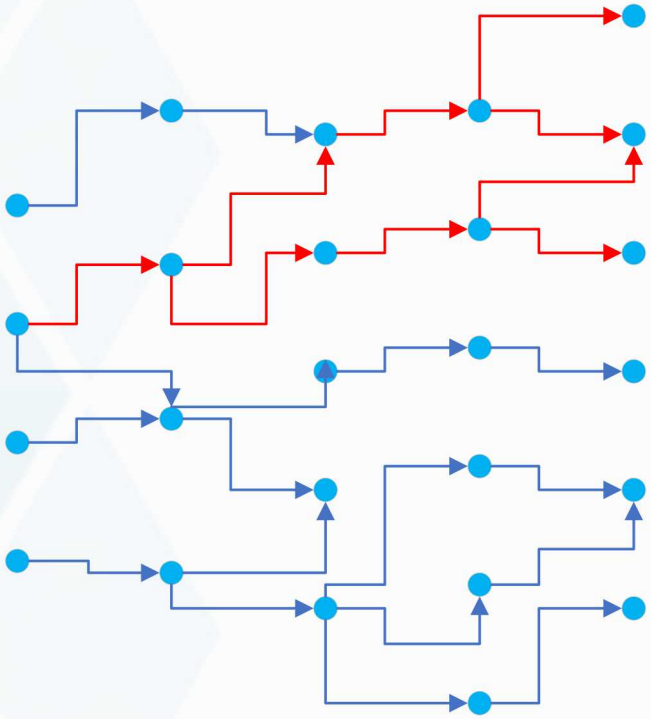
Attack sequences can only start at Initial nodes

Intermediate nodes can only be reached if at least one predecessor node is reached

Attack Graph Based Attack Model

A slightly more complicated example:

Multiple initial nodes possibly from multiple communication networks and/or multiple physical attackers



Relays at multiple substations can be compromised and allow attacker to open loads, generators, or lines

Combining kill chains into a single graph allows for analysis of efficient coordinated attacks

Worst-Case Attacker Model

$$\max_{x,y,u,v,w,z} \gamma(x,y,u,v,w,z)$$

$s.t.$

$$\sum_{e\in\mathcal{T}} D_e x_e \le B$$

$$x_{e'} \le \sum_{e\in\mathcal{T}_r} x_e$$

$$x_e \le y_r$$

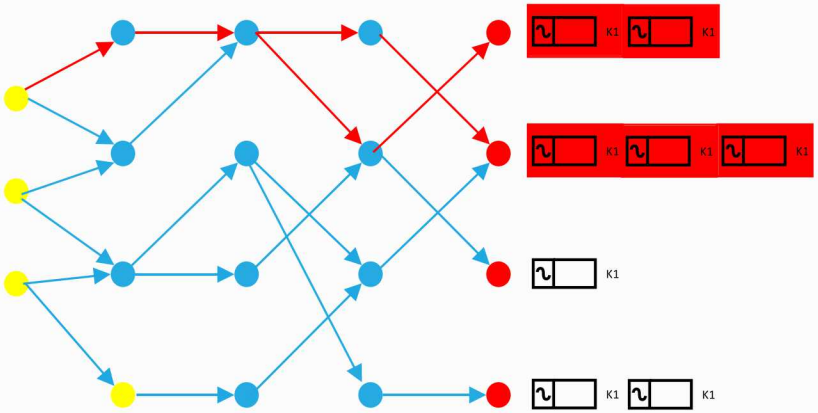
$$y_r \le \sum_{e\in\mathcal{T}_r} x_e$$

$$\sum_{r\in\mathcal{R}_l}(1-y_r) - |\mathcal{R}_l| + 1 \le u_l \le (1-y_r)$$

$$\sum_{r\in\mathcal{R}_k}(1-y_r) - |\mathcal{R}_k| + 1 \le v_k \le (1-y_r)$$

$$\sum_{r\in\mathcal{R}_g}(1-y_r) - |\mathcal{R}_g| + 1 \le w_g \le (1-y_r)$$

Attack Model



$$\gamma(x,y,u,v,w,z) = \min_{\theta,p,p^G,p^{L,S}} \sum_{b\in\mathcal{B}} p_b^{L,S}$$

$s.t.$

$$p_k = v_k B_k(\theta_{o(k)} - \theta_{d(k)} - \Theta_k)$$

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

$$-S_k^{\max} \le p_k \le S_k^{\max}$$

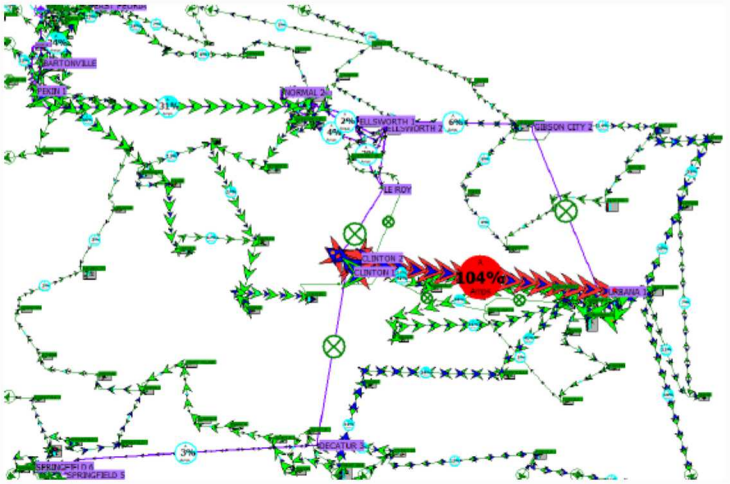
$$w_g P_g^{G,min} \le p_g^G \le w_g P_g^{G,max}$$

$$\sum_{l\in\mathcal{L}_b}(1-u_l)P_l^L \le p_b^{L,S} \le \sum_{l\in\mathcal{L}_b} P_l^L$$

$$-\pi \le \theta_b \le \pi$$

Damage Control

Optimal Power Flow



*Derived from synthetic data that does not represent actual grid: https://electricgrids.engr.tamu.edu/electric-grid-test-cases/activsg2000/

Worst-Case Interdiction Analysis of Large-Scale Electric Power Grids
Javier Salmeron-Kevin Wood-Ross Baldick - IEEE Transactions on Power Systems - 2009

Worst-Case Attacker Model Status

- Data Requirements from Emulytics and UQ
  - Attack graph
  - Edge weights: how hard is it to get from source node to destination node?
  - RTU mapping: if a relay is compromised, what grid components are opened?

- Solution Technique
  - Dualize defender problem and collapse into a single-level mixed-integer program (MIP)
  - Use commercial solvers

- Problem Difficulties
  - Models transformation leads to a MIP that can be difficult to solve due to numerical conditioning
  - New structure of reduced model can create computational difficulties

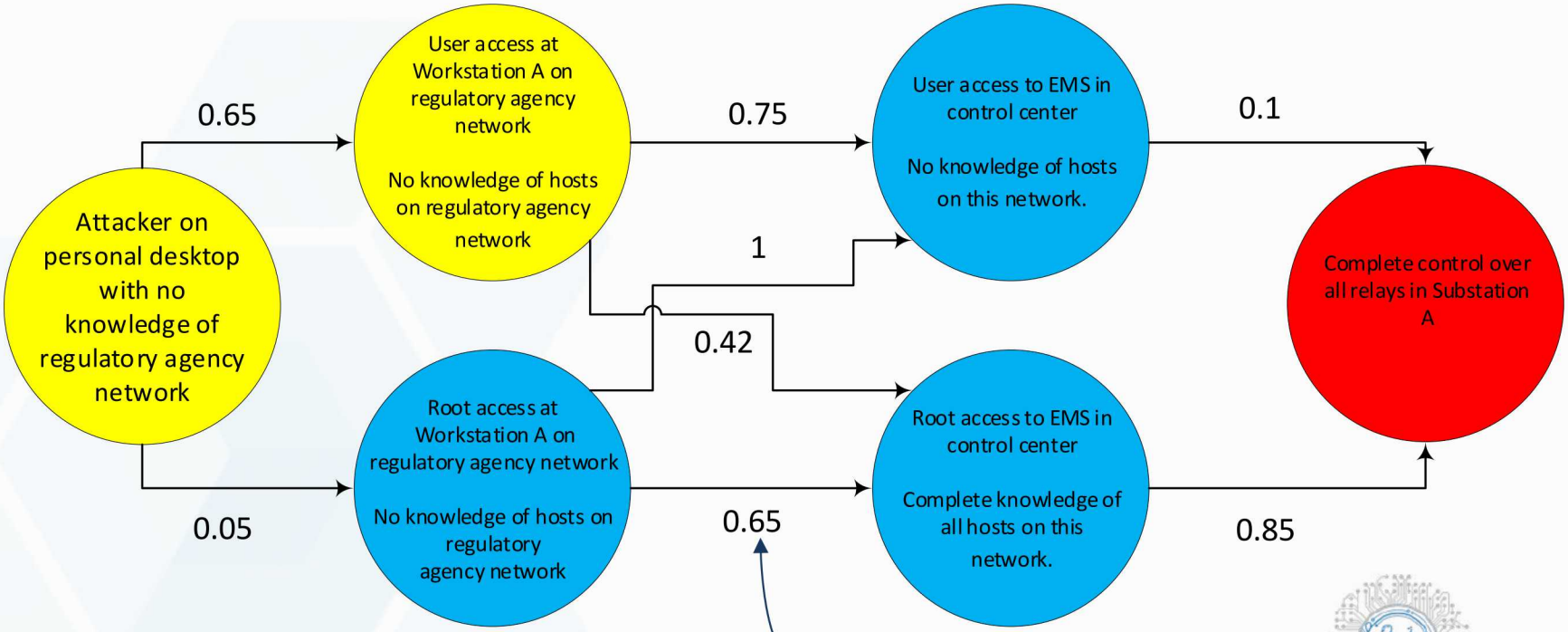
Comparison with GPLADD

- Only one stage per decision maker
- No attacker -> defender -> attacker -> defender -> attacker -> defender

- Our models explore a search space with exponentially many choices to make high-level decisions

- Future goal is to create synergy between optimization bilevel models and GPLADD

Stochastic Attack Graph



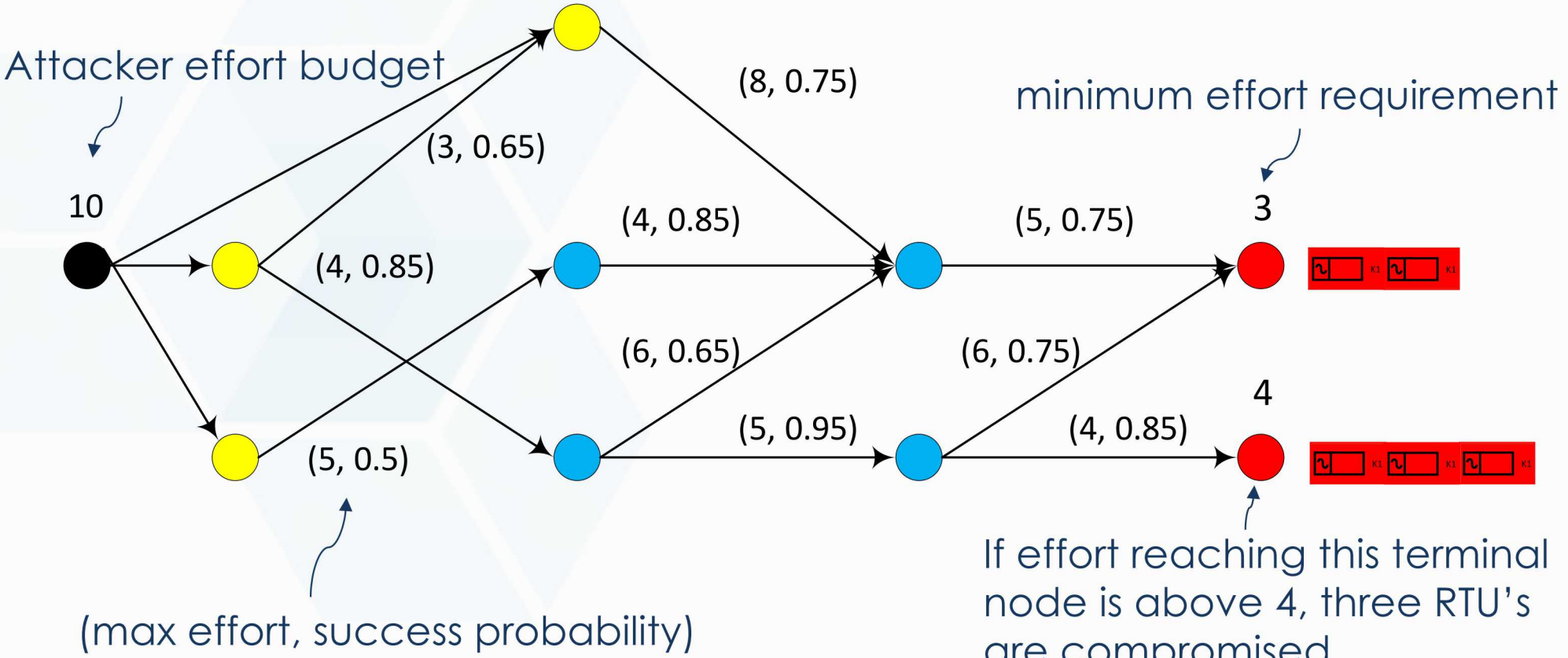
Now let's add edge probabilities to model difficulty in moving between nodes

EMULYTICS

DAKOTA

Stochastic Attack Graph Based Attack Model

- Attacker pushes "effort" through the network from black node to terminal nodes
- Edge probabilities cause effort leaking
- Total flow out = Total flow in after leaking
- Effort threshold is used to determine if RTU's are compromised



Feng Pan, William S. Charlton, and David P. Morton. A stochastic program for interdicting smuggled nuclear material. In Network Interdiction and Stochastic Integer Programming, volume 22, pages 1–19. Springer, 2003.

Stochastic Worst-Case Attack Model

$$\max_{a,z,\delta,u,v,w} \gamma(u,v,w)$$

$s.t.$

$$\sum_{e\in\mathcal{E}_{F(s)}} a_e = B$$

$$z_s = \sum_{e\in\mathcal{E}_{T(s)}} P_e^{\omega} a_e$$

$$\sum_{e\in\mathcal{E}_{F(s)}} a_e = \sum_{e\in\mathcal{E}_{T(s)}} P_e^{\omega} a_e$$

$$a_e \le u_e$$

$$t_s \delta_r \le z_s$$

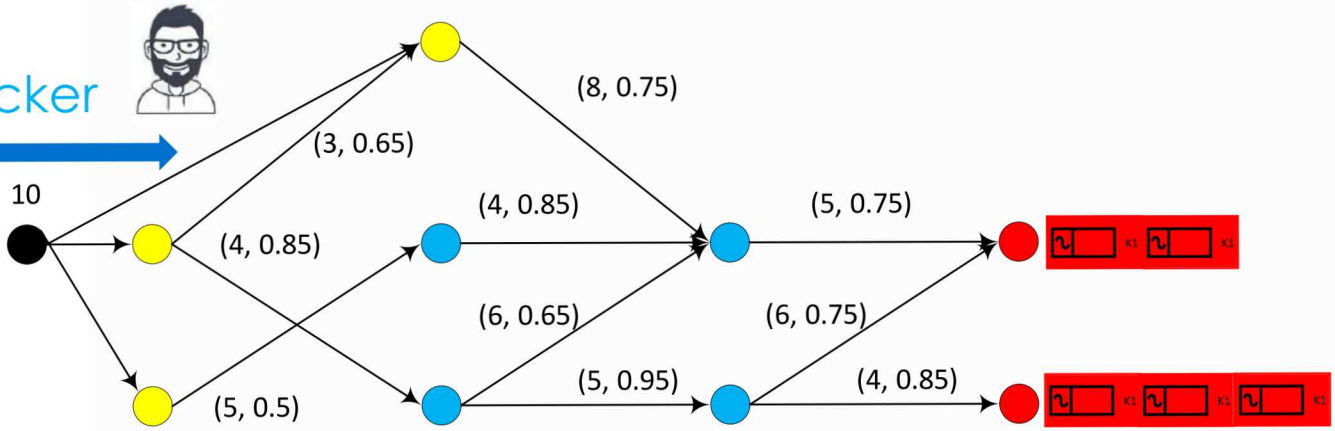
$$\sum_{r\in\mathcal{R}_l}(1-y_r) - |\mathcal{R}_l| + 1 \le u_l \le (1-y_r)$$

$$\sum_{r\in\mathcal{R}_k}(1-y_r) - |\mathcal{R}_k| + 1 \le v_k \le (1-y_r)$$

$$\sum_{r\in\mathcal{R}_g}(1-y_r) - |\mathcal{R}_g| + 1 \le w_g \le (1-y_r)$$

$$\gamma(u,v,w) = \min_{\theta,p,p^G,p^{L,S}} \sum_{b\in\mathcal{B}} p_b^{L,S}$$

Stochastic Attacker



$s.t.$

$$p_k = v_k B_k(\theta_{o(k)} - \theta_{d(k)} - \Theta_k)$$

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

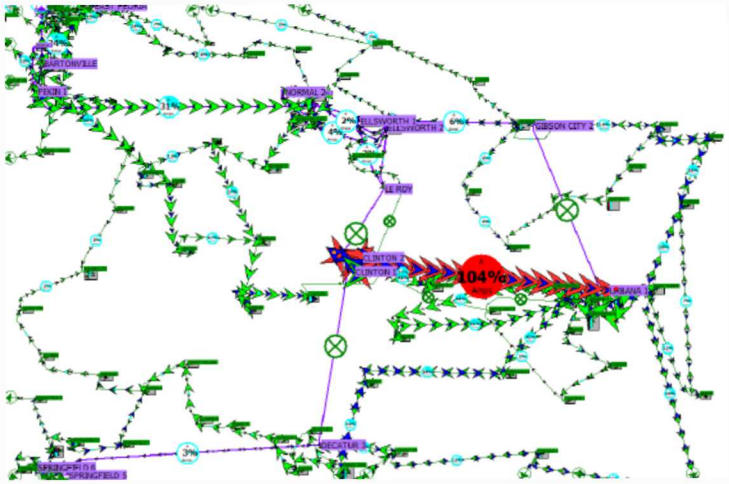
$$-S_k^{\max} \le p_k \le S_k^{\max}$$

$$w_g P_g^{G,min} \le p_g^G \le w_g P_g^{G,max}$$

$$\sum_{l\in\mathcal{L}_b}(1-u_l)P_l^L \le p_b^{L,S} \le \sum_{l\in\mathcal{L}_b} P_l^L$$

$$-\pi \le \theta_b \le \pi$$

Damage Control

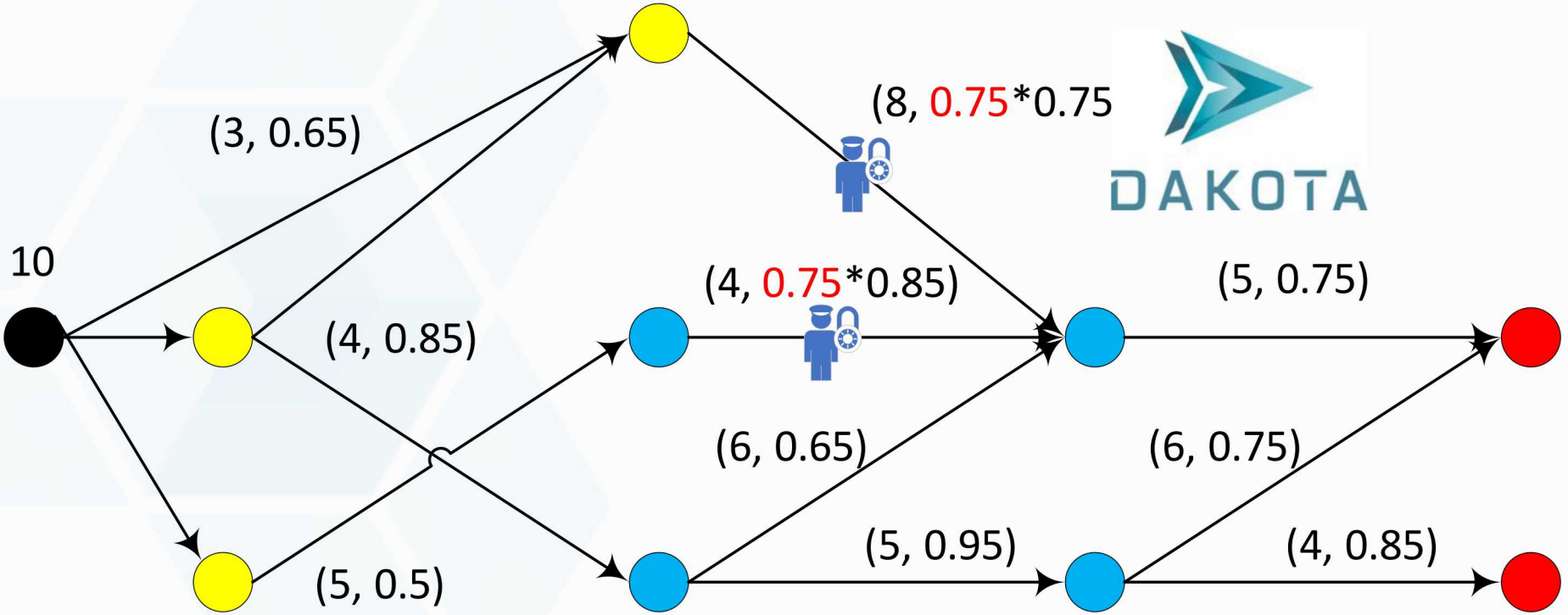


*Derived from synthetic data that does not represent actual grid: https://electricgrids.engr.tamu.edu/electric-grid-test-cases/activsg2000/

Stochastic Worst-Case Attacker Model Status

- Data Requirements from Emulytics and UQ
  - Attack graph
  - Edge probabilities: probability of not being detected when moving between nodes
  - RTU mapping: if a relay is compromised, what grid components are opened?
  - Attacker effort budget: how much total effort does the attacker have available to expend?
  - RTU effort: how much effort does each RTU need before the attacker controls it?
- Solution Technique
  - Dualize defender problem and collapse into a single-level mixed-integer program (MIP)
  - Use commercial solvers
- Problem Difficulties
  - Models transformation leads to a MIP that can be difficult to solve due to numerical conditioning
  - New structure of reduced model can create computational difficulties

Intrusion Detection System Placement



Intrusion Detection System Placement Model

$$\min_{x} \lambda(x)$$

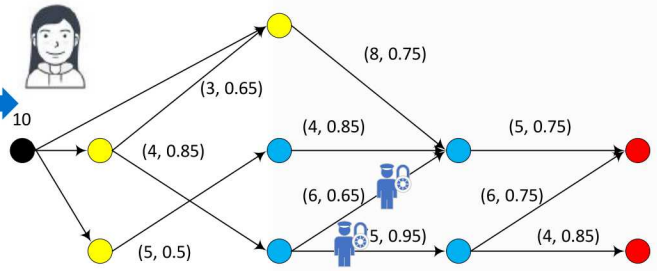
Network Designer

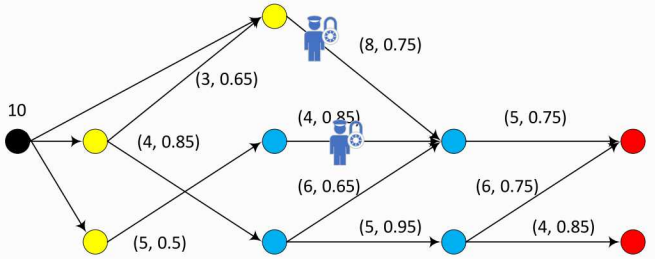
$$s.t.$$

$$\sum_{i \in \mathcal{I}} c_i x_i \leq B$$

$$\lambda(x) = \max_{a,z,\delta,u,v,w} \gamma(u,v,w)$$

EMULYTICS





$$s.t.$$

$$\sum_{e \in \mathcal{E}_{F(s)}} a_e = B$$

$$z_s = \sum_{e \in \mathcal{E}_{T(s)}} P_e^{\omega} a_e$$

$$\sum_{e \in \mathcal{E}_{F(s)}} a_e \leq \sum_{e \in \mathcal{E}_{T(s)}} a_e P_e^{\omega} \prod_{i \in \mathcal{I}_e} (Q_{i,e} x_i + (1 - x_i))$$

$$a_e \leq u_e$$

$$t_s \delta_r \leq z_s$$

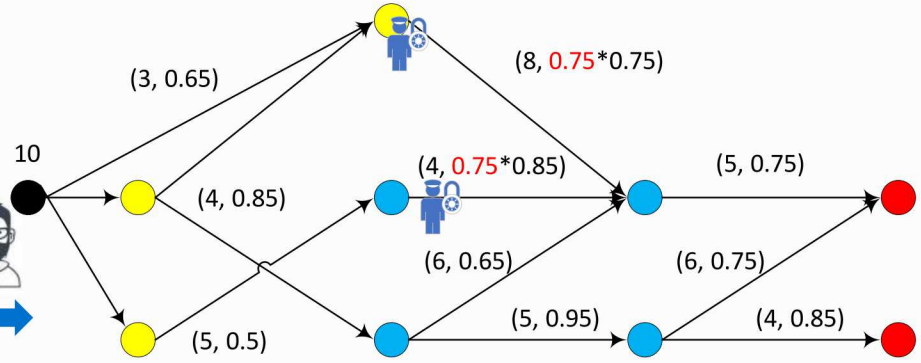
$$\sum_{r \in \mathcal{R}_l} (1 - y_r) - |\mathcal{R}_l| + 1 \leq u_l \leq (1 - y_r)$$

$$\sum_{r \in \mathcal{R}_k} (1 - y_r) - |\mathcal{R}_k| + 1 \leq v_k \leq (1 - y_r)$$

$$\sum_{r \in \mathcal{R}_g} (1 - y_r) - |\mathcal{R}_g| + 1 \leq w_g \leq (1 - y_r)$$

$$\gamma(u,v,w) = \min_{\theta, p, p^G, p^{L,S}} \sum_{b \in \mathcal{B}} p_b^{L,S}$$

Stochastic Attacker



$$s.t.$$

$$p_k = v_k B_k (\theta_{o(k)} - \theta_{d(k)} - \Theta_k)$$

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

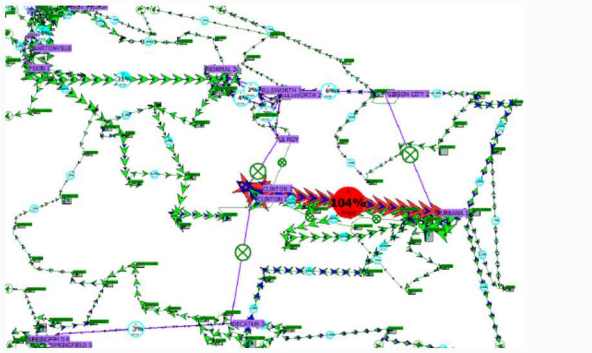
$$-S_k^{\max} \leq p_k \leq S_k^{\max}$$

$$w_g P_g^{G,min} \leq p_g^G \leq w_g P_g^{G,max}$$

$$\sum_{l \in \mathcal{L}_b} (1 - u_l) P_l^L \leq p_b^{L,S} \leq \sum_{l \in \mathcal{L}_b} P_l^L$$

$$-\pi \leq \theta_b \leq \pi$$

Damage Control



*Derived from synthetic data that does not represent actual grid: https://electricgrids.engr.tamu.edu/electric-grid-test-cases/activsg2000/

IDS Placement Model Status

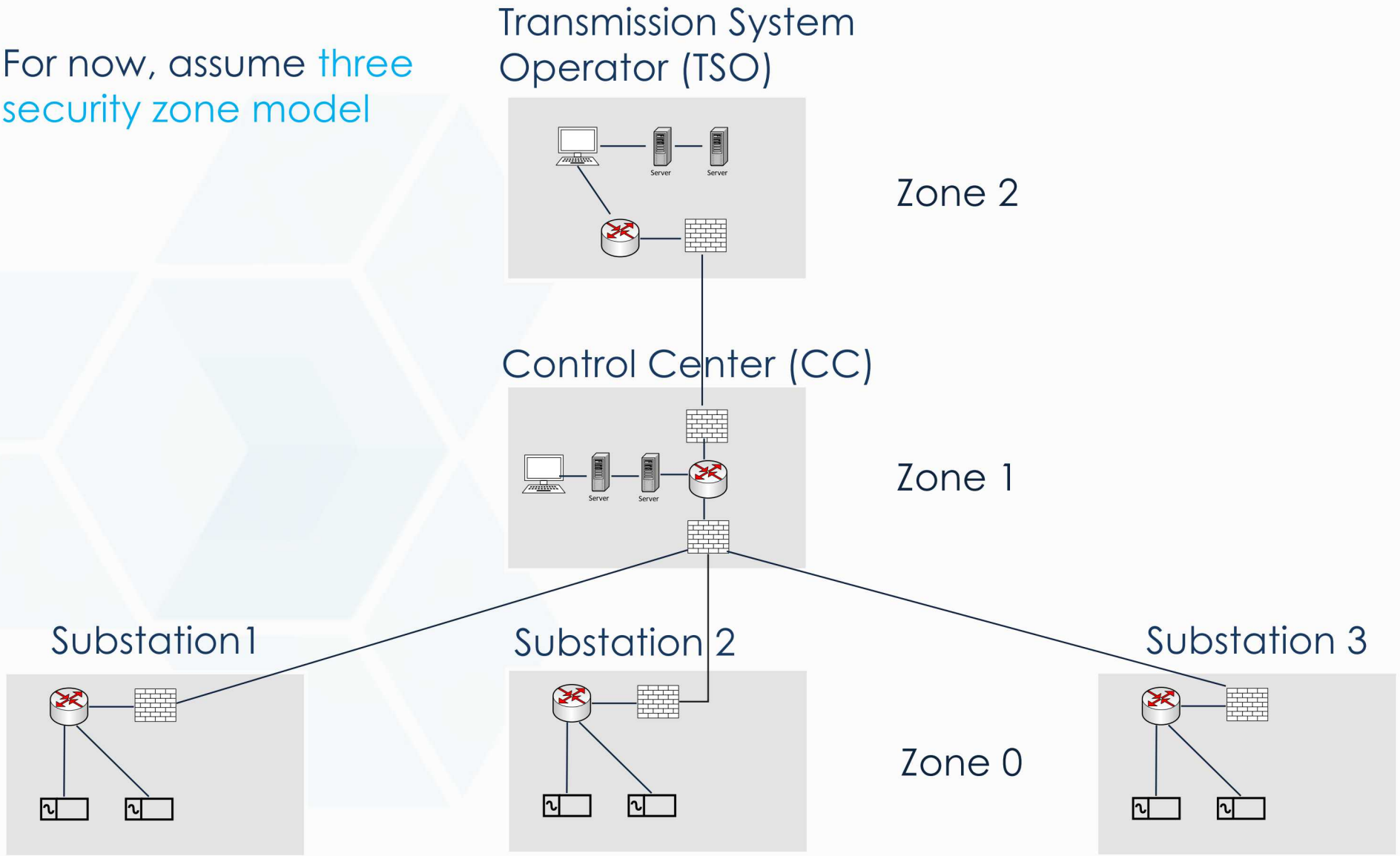
- Data Requirements from Emulytics and UQ
  - All data from stochastic worst-case scenario attack model
  - Sensor costs and budget
  - Sensor probability multipliers: if a sensor is purchased, which edge probabilities are affected and by how much are the probabilities decreased?

- Solution Technique
  - Dualize inner problem and reduce to a difficult mixed-integer bilevel program
  - Try MibS, Fischetti solver, and recent algorithms

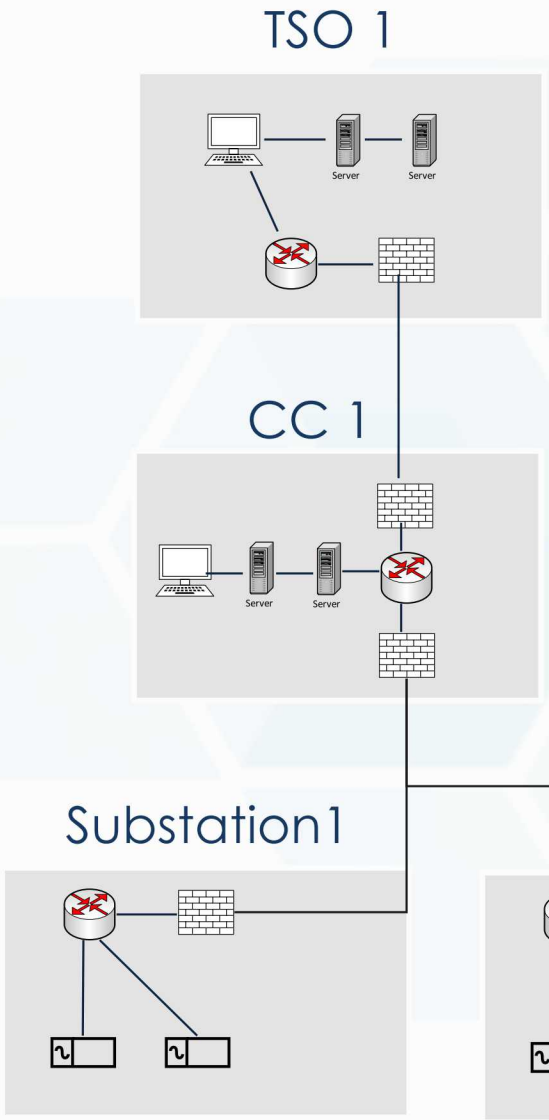
- Problem Difficulties
  - The resulting bilevel program cannot be reduced again into a single MIP
    - Leader has both continuous and discrete variables

Network Segmentation Problem

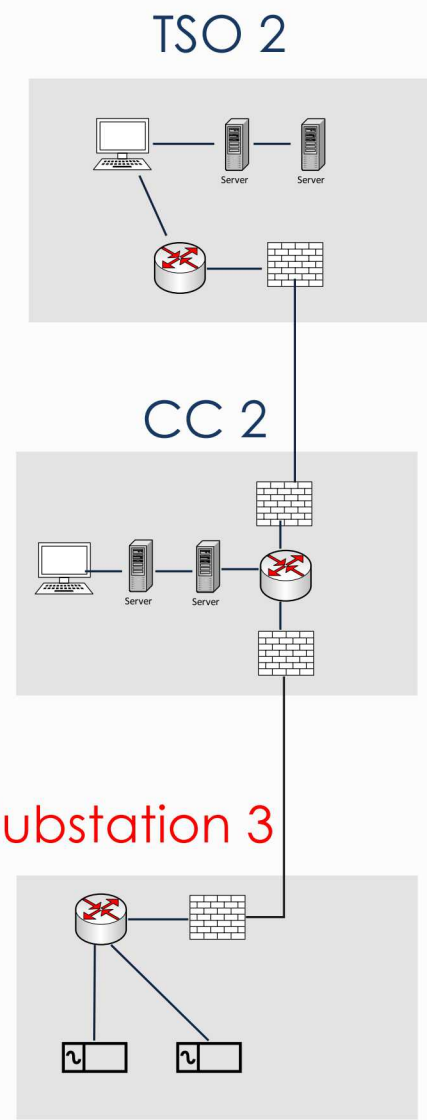
For now, assume three security zone model



Network Segmentation Problem



- The grid can be severely damaged when Substation 2 and Substation 3 are attacked together.
- Substation 1 and Substation 2 are configured so that the grid is fine if they are attacked together.



Network Segmentation Model

$$\min_{x,y} \gamma(x,y)$$

$$s.t.$$

$$\sum_{f\in\mathcal{F}-\{T\}} x_{r,f} = 1$$

$$\sum_{\{f>e\}} y_{e,f} = 1$$

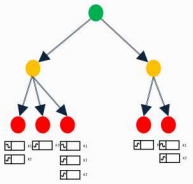
$$l_T = 3$$

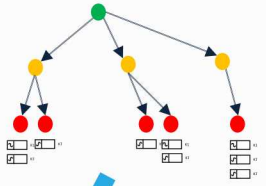
$$l_f \le 2(1-\sum_{r\in\mathcal{R}} x_{r,f})$$

$$l_f \ge y_{e,f}(l_e+1)$$

$$l_e \le y_{e,f}(l_f-1)+2(1-y_{e,f})$$

Network Segmentation





EMULYTICS

$$\gamma(x,y) = \max_{z,\delta} \lambda(u,v,w)$$

$$\sum_{e\in\mathcal{F}} z_e \le B$$

$$z_e \le \sum_{f>e} y_{e,f} z_f + y_{e,T}$$

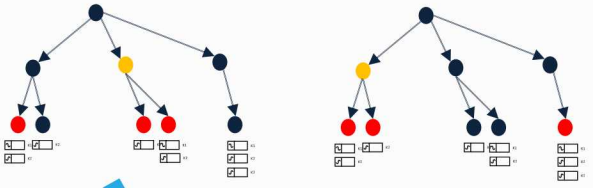
$$\delta_r = \sum_{e\in\mathcal{F}} x_{r,e} z_e$$

$$\sum_{r\in\mathcal{R}_k} (1-\delta_r) - |\mathcal{R}_k| + 1 \le v_k \le (1-\delta_r),$$

$$\sum_{r\in\mathcal{R}_l} (1-\delta_r) - |\mathcal{R}_l| + 1 \le u_l \le (1-\delta_r),$$

$$\sum_{r\in\mathcal{R}_l} (1-\delta_r) - |\mathcal{R}_l| + 1 \le w_g \le (1-\delta_r),$$

Attack Model



$$\lambda(u,v,w) = \min_{\theta,p,p^G,p^{L,S}} \sum_{b\in\mathcal{B}} p_b^{L,S}$$

$$s.t.$$

$$p_k = v_k B_k(\theta_{o(k)} - \theta_{d(k)} - \Theta_k)$$

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

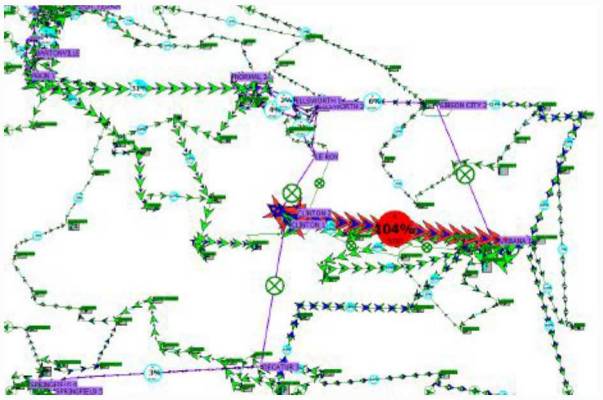
$$-S_k^{\max} \le p_k \le S_k^{\max}$$

$$w_g P_g^{G,max} \le p_g^G \le w_g P_g^{G,max}$$

$$\sum_{l\in\mathcal{L}_b} (1-u_l) P_l^L \le p_b^{L,S} \le \sum_{l\in\mathcal{L}_b} P_l^L$$

$$-\pi \le \theta_b \le \pi$$

Damage Control



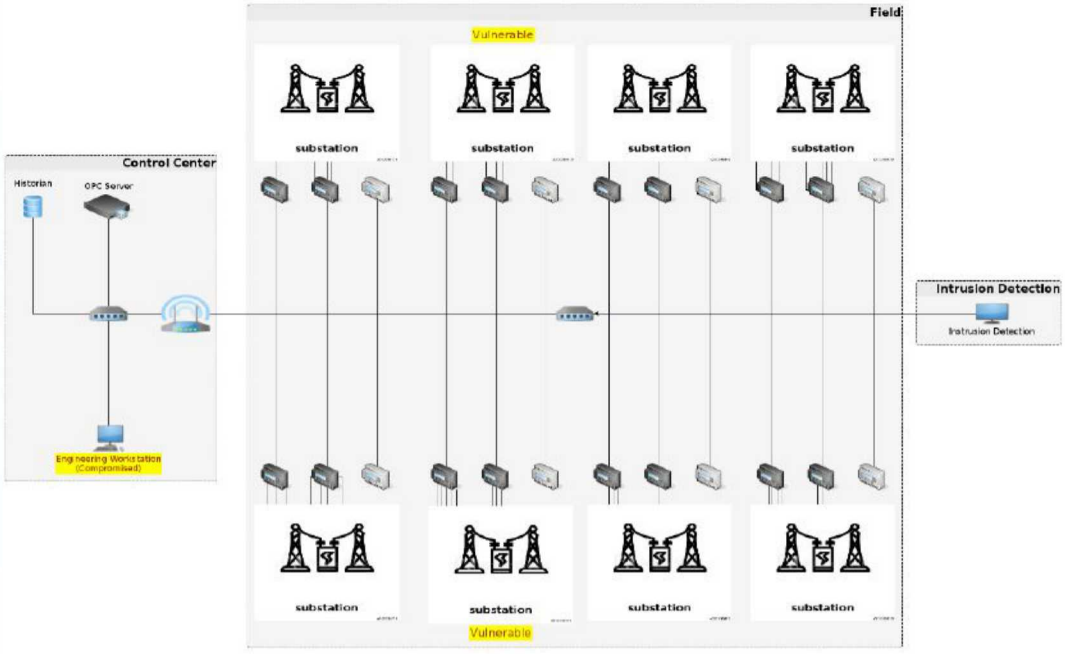
*Derived from synthetic data that does not represent actual grid: https://electricgrids.engr.tamu.edu/electric-grid-test-cases/activsg2000/

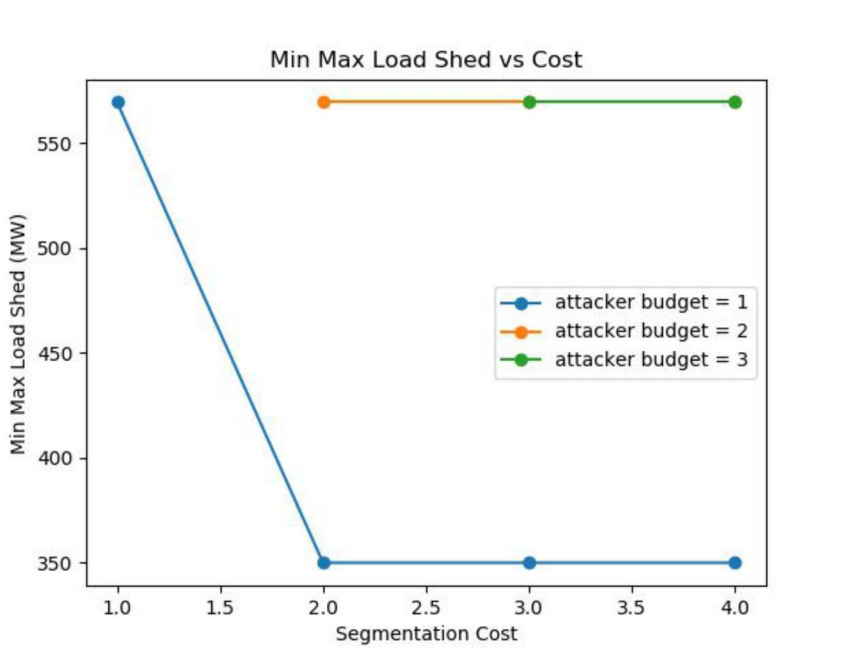
Network Segmentation Model Status

- Data Requirements
  - Network segmentation budget
  - Attacker budget
- Solution Technique
  - Dualize inner problem and reduce to a difficult mixed-integer bilevel program
  - Try MibS, Fischetti solver, and recent algorithms
- Problem Difficulties
  - The resulting bilevel program cannot be reduced again into a single MIP
    - Leader has both continuous and discrete variables
- Advantage
  - Data requirements are minimal
  - Unlike other models, this one does not require detailed Emulytics/UQ data.

Network Segmentation Results

The current 4-RTU exemplar is small enough to perform optimization through complete enumeration of all choices for designer, attacker, and defender





Future Extensions of Network Segmentation

- Network segmentation pricing
  - Assign a cost to each subnet that depends on security zone
  - Use a budget to limit the overall cost of network segmentation

- If necessary, add subnet detail so that a subnet is more than just a node. Preferably don't since this model requires minimal SME data.
  - Use caution when adding model detail. We must remember that these bilevel models are incredibly difficult to solve

- Add automated network segmentation into Emulytics

SCORCH

Future Optimization Problems

- Network scanning optimization
  - Use optimization to pick optimal network scanning parameters
    - Number of nodes to scan in parallel
    - Probe delay
    - Number of retries
- Optimize over RTU connections to loads, lines, and generators to suggest more resilient cyber physical configurations
- Provide Emulytics team higher-fidelity power flow capability

Conclusion

- Our bilevel and trilevel models are driving discussions on what type of optimization problems we should formulate
  - What kind of data can we expect to get from Emulytics/UQ experiments?
  - Are the questions that these models address interesting to the rest of the team?
  - If V&V effort indicates that model fidelity is an issue, we can add detail and reiterate.
- After year 1, we have a promising suite of tools for trying to solve our difficult models
- Santanu Dey from Georgia Tech is excited about our models and is ready to use his expertise to help us
- These models have huge publication potential
  - Easily one publication per model

LDRD

Laboratory Directed Research and Development

SECURE Summary

Ali Pinar, PI

Zach Benz, PM

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.

Overview of the Exemplar Study/Workflow

Threat Model

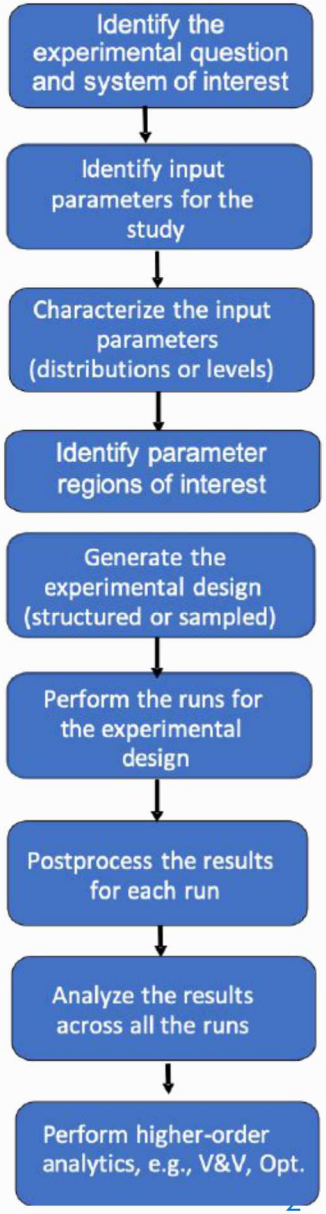
- Crashoverride on a single ICS
- Focus on part of the attack Reconnaissance
- Attacker needs to act quickly
- Attacker tries to locate RTUs using *nmap*
- Defender tries to detect such searches using *snort*
- Parameter ranges set for a fast strike attack

Attack Effect on Resources

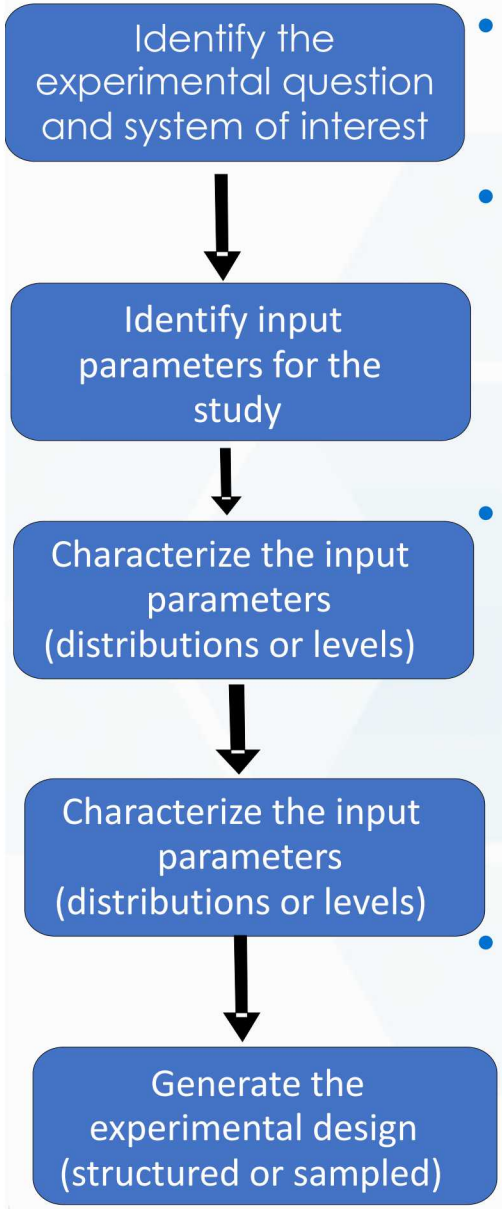
- A representation of the a region of the Texas Grid
- Flat cyber network
- Controls 8 RTUs
- Build an emulation model of the system
- Run the emulation many times to cover the parameter space
- Build models for impact on cyber
- Validate models using emulation

Consequence Prediction

- Quantify impact on the power grid based on loss of load
- Investigate how a sophisticate adversary can use this attack in an optimal way
- Provide feedback to previous steps about sensitive parameters regions



Understand the Threat



- Question: how do we improve our resilience against crashoverride on the power grid?
- Focus on the reconnaissance step of the attack chain
  - Attack tool: nmap, to locate RTUs in the network
  - Defense tool: Snort to detect network scans
- Setting:
  - An adversary can get to the system through a phishing attack
  - Once in the system, it has a small time-window to operate.
  - Restrict nmap parameters to this space
- Workflow steps:
  - Model the system
    - Network architecture, parameters, RTU placement

Impact of the threat on the cyber system

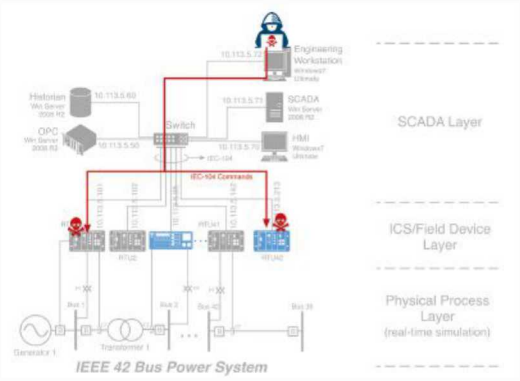
Perform the runs for the experimental design

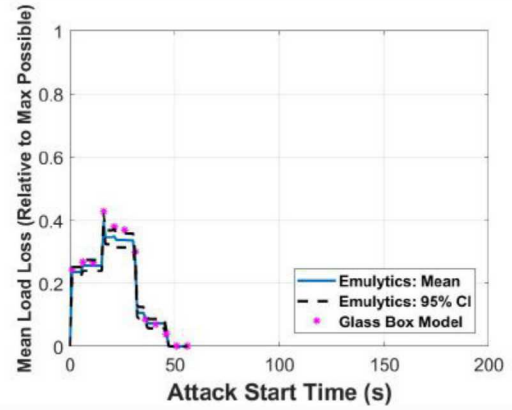
Postprocess the results for each run

Analyze the results across all the runs

Perform higher-order analytics, e.g., V&V, Opt.

- Execute the model to search over the parameter space to propagate uncertainties to build distribution of the attack impact
  - How many RTUs are likely to be compromised?
  - What does the tail look like?
  - Are there correlations between pairs of RTUs being compromised?
- Verify, and validate
  - the original emulation model
  - and the abstractions we build for higher order analytics
- Answer higher questions within the cyber system
  - What is the best way to attack?
  - What is the best way to defend?
  - What are alternative ways to defend?

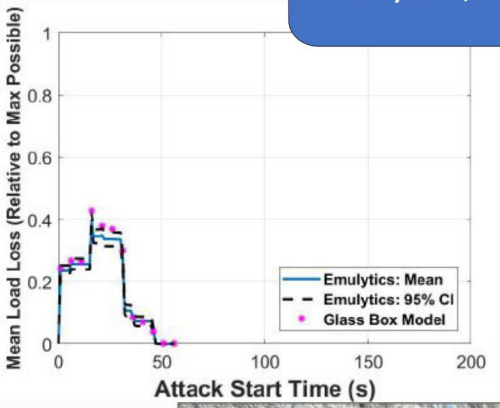


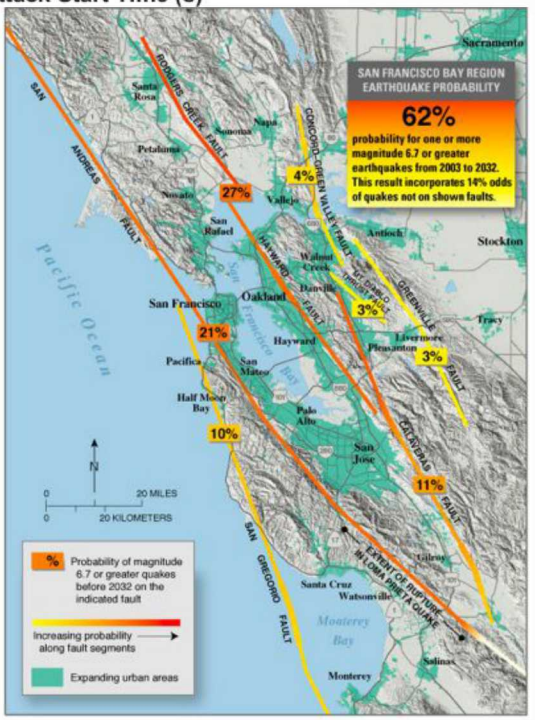


Tie between cyber and physical systems

- Given the models of the attack on the cyber system,
  - What are the consequences on the physical system?
  - How can the adversary use the attack for maximum damage?
  - How do we operate on the physical system with cyber awareness?
  - N-k security has a new meaning now.
- Feedback to the earlier steps
  - How do we build cyber systems for better physical resilience?
    - Network segmentation
    - Judicious intrusion Detection Systems
  - What are the sensitive model parameters that need to be captured with higher fidelity?

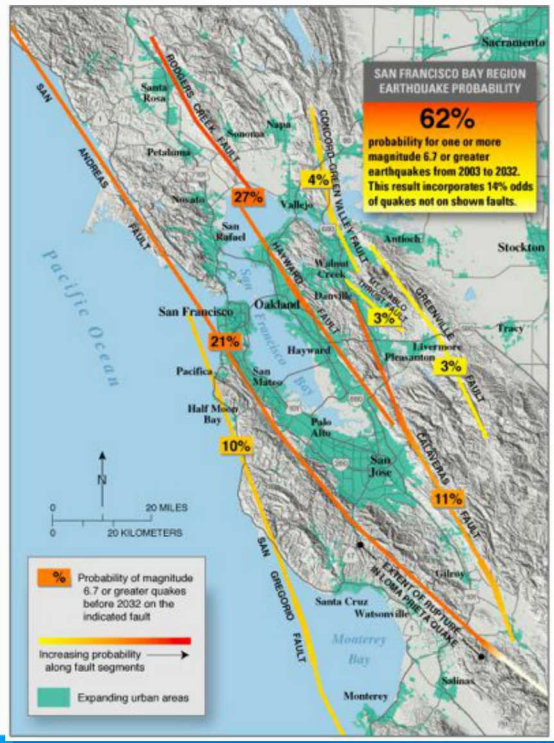
Perform higher-order analytics, e.g., V&V, Opt.





Cyber-aware resilience and Consequence-aware cyber defense

- Cyber attacks lead to correlated physical failures.
  - What is a cyber fault line?
- How do we prioritize our defenses?
- How effective are our defenses?
  - Is one solution quantifiable better than another?
- How do we improve cyber-systems for better resilience?
- How do we operate on physical systems in a cyber threat-informed way?



Research Plan (Overview)

Year 1: Integration and Algorithmic Exploration

- Surveys; apply present capabilities; integration (tools and ideas); initial results for new ideas; fine-tuned problem definitions
- Exemplar 1: Single operating authority; flat SCADA/RTU network;
- Products: Prototype implementations; papers on early results; integrated experimental environment

Year 2: Algorithm Development

- Deep dive into algorithmic research; testing at scale/complexity; research software; initial demonstration of new, joint capabilities
- Exemplar 2: Regional; SCADA/RTU network; multiple ICS networks
- Products: SECUREtk 0.1(internal use); Algorithm publications

Year 3: Demonstrate Capability

- Pushing the boundaries of tools; Reporting results; demonstration of capabilities; research software to tools;
- Exemplar 3: Western Grid; IT/SCADA/RTU network; multiple subnets & services
- Products: SECUREtk 1.0 (sharable with research partners); Integration publications

FY20 Plan: Integrated Overview

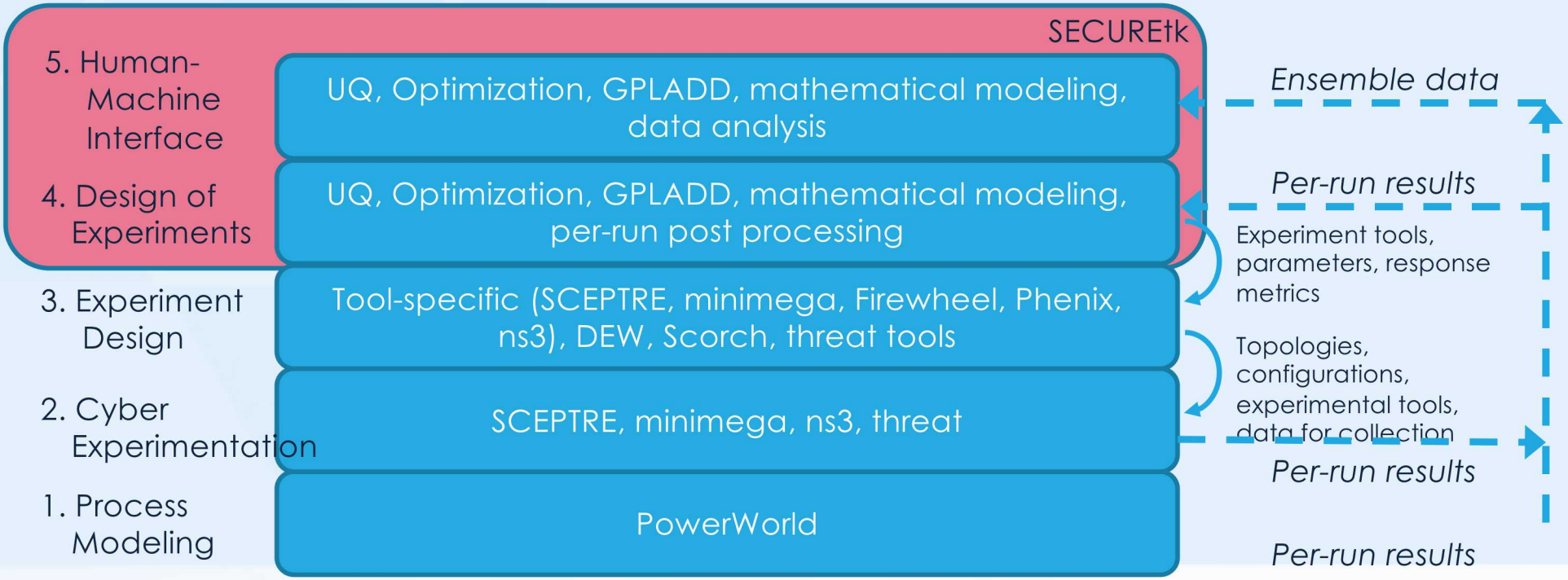
- **Scaling up the Exemplar**
  - More complex SCADA networks; Multiple ICS networks;
  - More complicated questions; restructuring the network; tailored defenses
  - Regional attacks: attacks on multiple ICS
  - Detailed models for higher dimensions
- **Leveraging Exemplars to Showcase Use Cases of SECUREtk**
  - Risk management through stochastic adversarial optimization
  - Design of Emulytics experiments with analytical methods
  - Constructing confidence intervals under high response variability
  - Hypothesis testing to guide experimental design
- **Quantifying V&V as part of the Cyber Experimental Process**
  - Conducting V&V in the context of problem
  - Requirements analysis to assess well-posedness of cyber models
  - Extending methods to address boundary conditions and estimating tail probabilities
- **Integration of Threat Characterization to Experiment Ensembles**
  - Pruning meaningless and low-consequence attack spaces prior to running Emulytics experiments
  - Identifying optimal mitigation strategies that deter or evolve the threat space

What is SECUREtk?

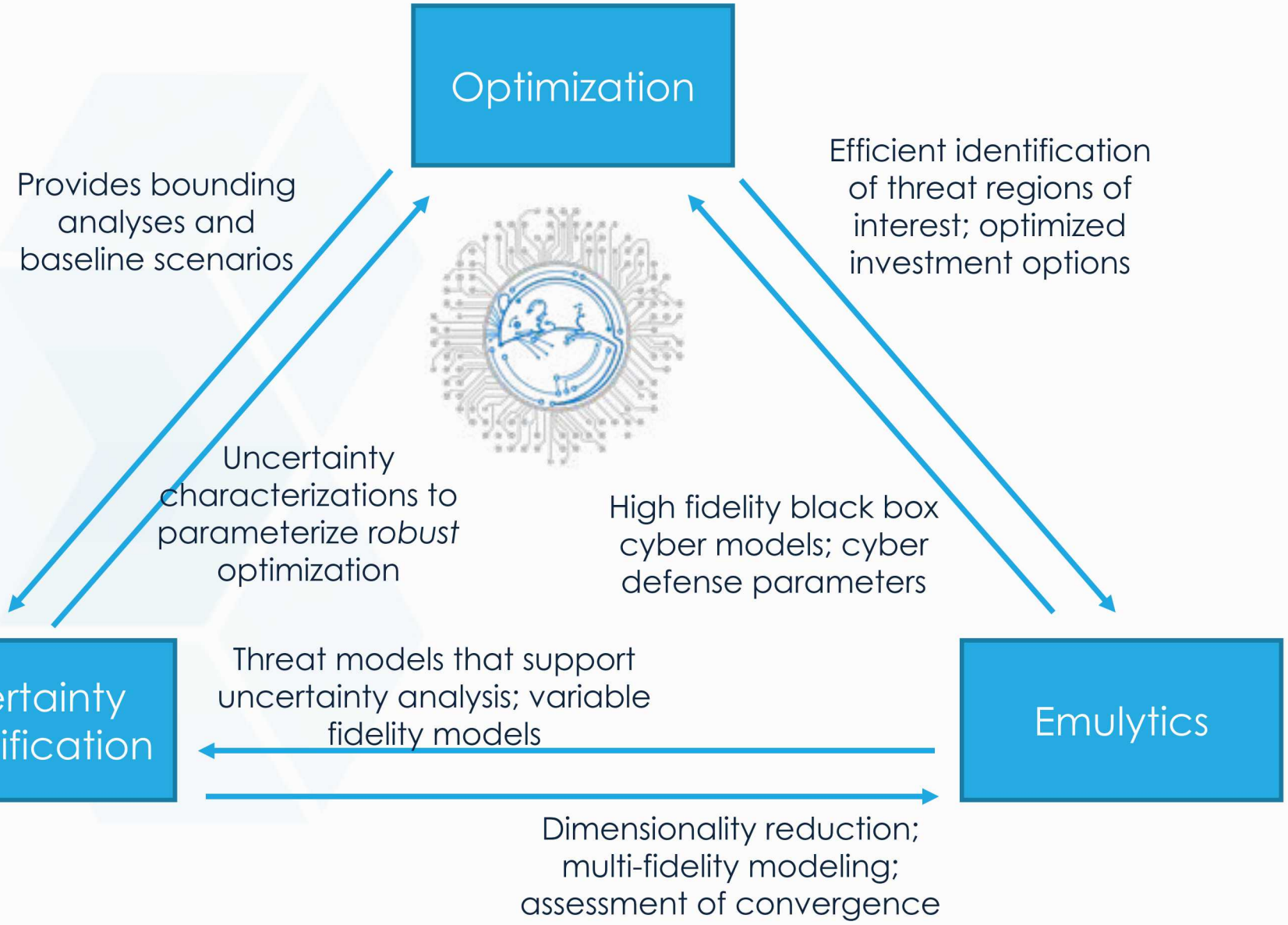
- SECUREtk is the ultimate goal, but it is not the top priority for now.
  - Not a driver for basic research, but driven by basic research
  - Currently developing building blocks
- Basic research is complemented by software development
  - E.g. Bi-level solvers for optimization
- It will be a collection of tools
  - Minimega, Firewheel, DAKOTA, Pyomo, etc.
  - Enables external contributions, flexibility for various systems
- Developing software for common needs
  - Scorch for many emulations with varying parameters
- FY20 Plan: develop a better understanding of the user profile and algorithmic tools
  - Details on Zach's talk on Wednesday morning
  - Let's talk about what to include on the wrap up discussion:

Cyber Experimental Software Stack

EMULYTICS



Research Element Dependencies



March EAB feedback items

CA
- Community awareness
  - "Future presentations should explicitly acknowledge related research and articulate how SECURE is going beyond it. "

RI
- Risks identification
  - "explicitly identify risks associated with the project and develop strategies to mitigate them "

TC
- Threat characterization
  - "it was not clear to the board how the threat characterization work contributed to the overall goals of the project "

ST
- SECUREtk

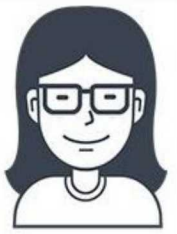
DE
- Domain expertise
  - "given that the exemplar is the power grid, the EAB did not see sufficient evidence that the team has the required domain expertise to create realistic scenarios that will answer meaningful questions "

PA
- Project architecture
  - "The board suggests that the team map out a project "architecture" that shows how tasks are connected … [and]the team needs to clearly define what comprises success and stake out integration activities to be accomplished throughout each year of the project"

CE
- Customer engagement
  - "it was not apparent either who the specific customers will be for SECURE's output or that the research plan is appropriately addressing medium- to long-term customer challenges "

PD
- People development
  - "The EAB was unclear on SNL's development / promotion of talent and expertise in cybersecurity"

Who can benefit from SECUREtk?



Adison the Engineer, IT decision maker

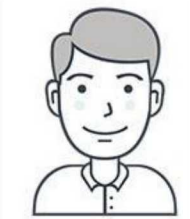
"Will deploying this cybersecurity solution have meaningful impact?"

- Her team offers different opinions about the potential benefits of the proposed solution and its impact on productivity
- She needs a thorough cost/benefit analysis to base her decision on

Captain Howard, DoD, high-consequence systems

"Can we credibly assess system performance under various threat scenarios?"

- He is in charge of a high-consequence system
- He trusts his red team, but the stakes are too high; the system is too complex; and time is too short



Leon the PM, capability steward

"What are the gaps in our capability roadmap to focus on to maximize impact? "

- He controls a budget that is too small; needs to prioritize
- Many conflicting expert opinions; system is too complex for the answers to be simple

Approximate Control Variate

M low-fidelity models with unknown expected value

For complex engineering models the **expected values of the M low-fidelity models are unknown** *a priori*

- Let's consider N_i LF evaluations: $N_i = \lfloor r_i N \rfloor$

The generic **Approximate Control Variate** is defined as

$$\tilde{Q}\left(\underline{\alpha}, \underline{\mathbf{z}}\right) = \hat{\mathbf{Q}}(\mathbf{z}) + \sum_{\mathbf{i=1}}^{\mathbf{M}} \alpha_{\mathbf{i}} \Delta_i(\mathbf{z}_i)$$

The **optimal weights and variance** can be obtained as

$$\underline{\alpha}^{\mathrm{ACV}} = -\mathbb{C}ov\left[\underline{\Delta}, \underline{\Delta}\right]^{-1} \mathbb{C}ov\left[\underline{\Delta}, \hat{Q}\right]$$

$$\mathbb{V}ar\left(\tilde{Q}\left(\underline{\alpha}^{ACV}\right)\right) = \mathbb{V}\mathbf{ar}\left(\hat{\mathbf{Q}}\right)\left(\mathbf{1} - \mathbf{R}^{\mathbf{2}}_{\mathrm{ACV}}\right).$$

NOTES:

1. For a single low-fidelity model: $R^2_{ACV-1} = \frac{\mathbf{r_1 - 1}}{\mathbf{r_1}} \rho_1^2$
2. We can build provably optimal estimators: $\rho_1^2 \leq R^2_{ACV} \leq R^2_{OCV}$