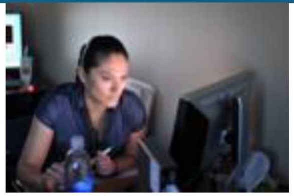
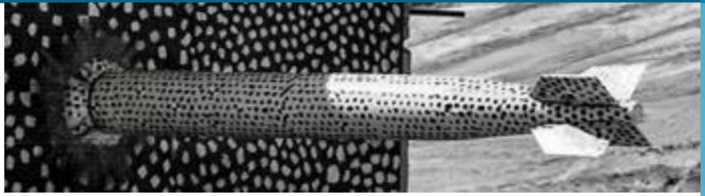
This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

Sandia National Laboratories

SAND2019-12762C

Uncertainty Quantification in Cyber Emulation





PRESENTED BY

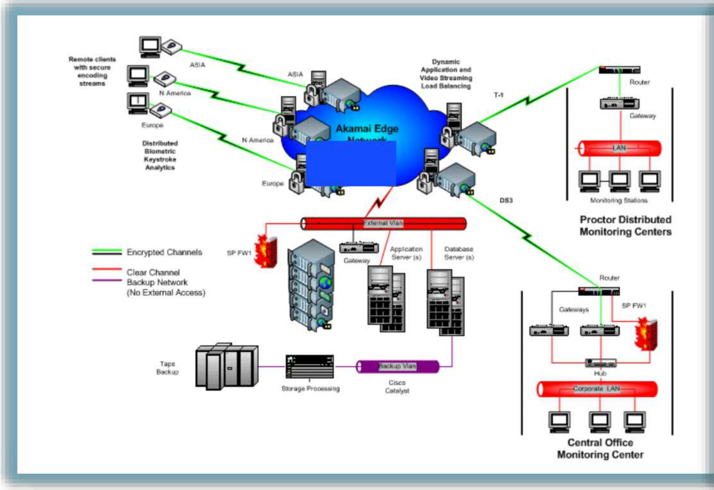
Laura Swiler

Gianluca Geraci, Bert Debusschere, Jonathan Crussell

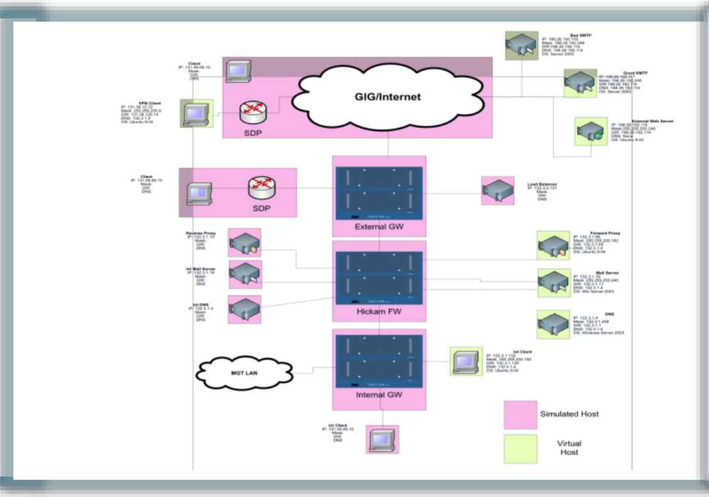
INFORMS Annual Meeting, Oct. 22, 2019

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.

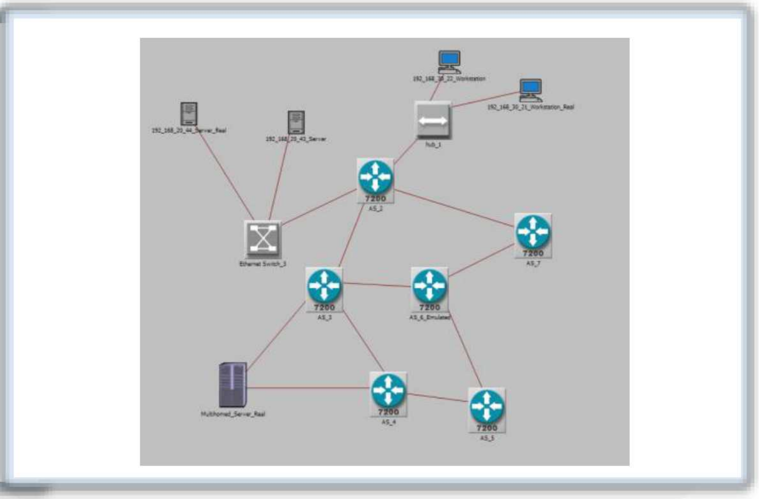
Emulation + Analytics = Emulytics


Remote clients with secure encoding streams
ASIA
N America
Europe
Distributed Biometric Keystroke Analytics
Dynamic Application and Video Streaming Load Balancing
Akamai Edge Network
T-1
DS3
Router
Gateway
LAN
Monitoring Stations
Proctor Distributed Monitoring Centers
Encrypted Channels
Clear Channel Backup Network (No External Access)
SP FW1
External Vlan
Gateway
Application Server (s)
Database Server (s)
Tape Backup
Storage Processing
Backup Vlan
Cisco Catalyst
Corporate LAN
Hub
Central Office Monitoring Center

ACTUAL SYSTEM



VIRTUALIZED TESTBED



SIMULATION TESTBED

Interoperability in a single experiment

LIVE ← Increase Realism

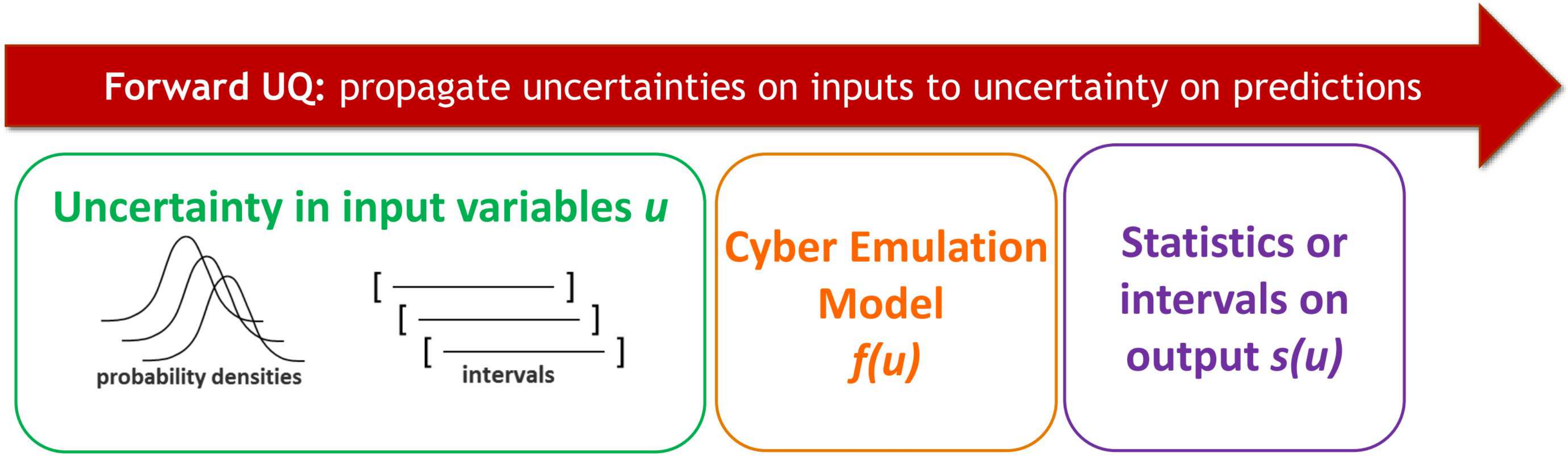
Decrease Cost, Decrease Time → SIMULATED

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

SECURE is integrating the mathematics of uncertainty quantification with emulytics to improve cyber experimentation.

What is Uncertainty Quantification?

- **Uncertainty Quantification** (UQ) is the process of characterizing all uncertainties that could affect the results of the cyber experimental runs.
- Once the uncertainties are identified and characterized as “input uncertainties”, they are propagated (e.g. mapped) through the experiment to obtain uncertainties on the results (“output uncertainties”).


Forward UQ: propagate uncertainties on inputs to uncertainty on predictions
Uncertainty in input variables *u*
probability densities
intervals
Cyber Emulation Model *f(u)*
Statistics or intervals on output *s(u)*

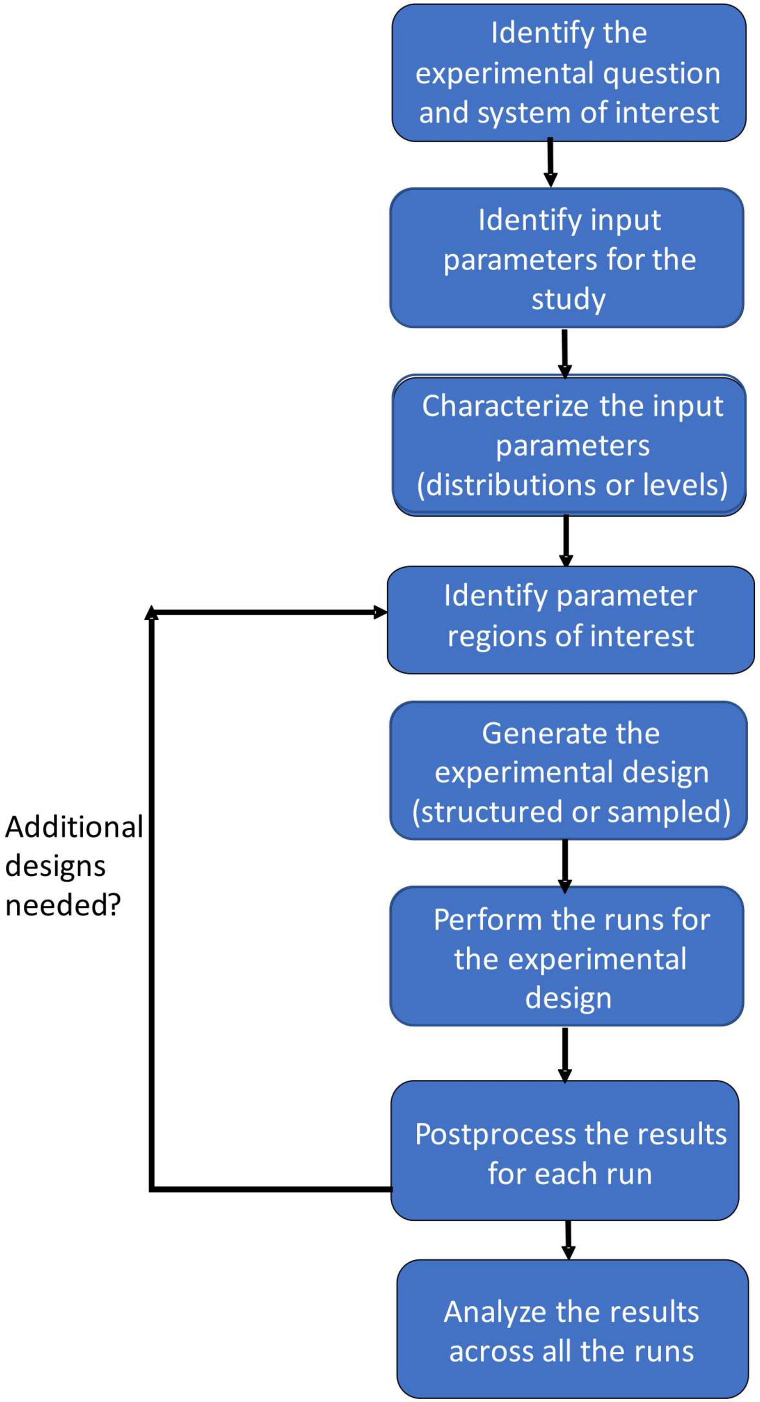
What is Experimental Design?

- **Experiment design** involves the selection of a set of experimental parameter settings at which one will run the cyber experimental model (e.g. various choices of number of cores per machine, protocols and environment settings, packet size of traffic, bandwidth of links, etc.)
- The goal of experimental design is to generate an "ensemble" of runs, where each run itself may involve replicate runs or not.
- Experiments are often tailored to address a specific question, such as: does the packet size matter when looking at packet response time on a network with sufficient bandwidth?
- There are a variety of design criteria.
  - Some experiments are space-filling (Monte-Carlo sampling and variations)
  - Some designs optimize some property of the run matrix, X.
  - A D-optimal design minimizes the determinant of the information matrix $[X^T X]^{-1}$ which results in maximizing the information content of the parameter values. A G-optimal design minimizes the maximum variance of the predicted values from a regression fit built on the dataset X.

There is significant overlap between UQ and experimental design. Both involve exploring the parameter space. UQ is more focused on distributional mapping.

Typical Workflow

- Think about the advantages of a structured vs. random sample design
- Think about the question you want to address
- Computational cost is a big factor


Identify the experimental question and system of interest
Identify input parameters for the study
Characterize the input parameters (distributions or levels)
Identify parameter regions of interest
Generate the experimental design (structured or sampled)
Perform the runs for the experimental design
Postprocess the results for each run
Analyze the results across all the runs
Additional designs needed?

More definitions

- **Sensitivity analysis** (SA) is the process of identifying the most significant factors or variables affecting the uncertainty of the cyber model predictions

- Over the past few decades, the computational simulation community has developed a strong emphasis on **Verification and Validation** activities to build credibility in scientific computing. A study by the National Research Council at the National Academies issued a report outlining the mathematical and statistical foundations of V&V and UQ as primary activities supporting the reliability of computational models [1]. We take as definitions those outlined in [2]:

❑ *Verification* is the process of assessing software correctness and numerical accuracy of the solution to a given mathematical model.

❑ *Validation* is the process of assessing the physical accuracy of a mathematical model based on comparisons between computational results and experimental data.

1. National Research Council. Assessing the Reliability of Complex Models: Mathematical and Statistical Foundations of Verification, Validation, and Uncertainty Quantification. Washington, DC: The National Academies Press, 2012. https://doi.org/10.17226/13395.
2. Oberkampf, W.L. and C.J. Roy. Verification and Validation in Scientific Computing. Cambridge University Press, 2010.

A few more things

- **Factorial design:** A factorial design is an experimental design that samples the full combination of all parameters.
  - Thus, if there were 3 parameters and each had 5 allowable values or levels, a full factorial design would involve 5*5*5 = 125 runs.
  - A fractional factorial design only involves a subset of the full factorial.
  - A full factorial design is an orthogonal array and can be used for main effects analysis

- **Replicates.** A replicate refers to running the same set of experimental settings multiple times to see how the response varies with-in that setting. A replicate can also be called an iterate.

Dakota: Toolkit of Sensitivity analysis, UQ, and experimental design methods to drive external codes

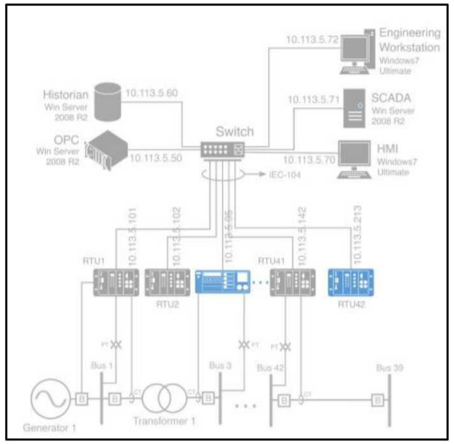
DAKOTA

Inputs

System Specification
- Devices
- Configuration
- Topology
- Connectivity
- Physical Processes

Threat Scenario:
- Actual malware
- Specify threat effect (e.g., kill RTU1)
- Red Team

Emulation Platform: VMs, HITL, Simulation

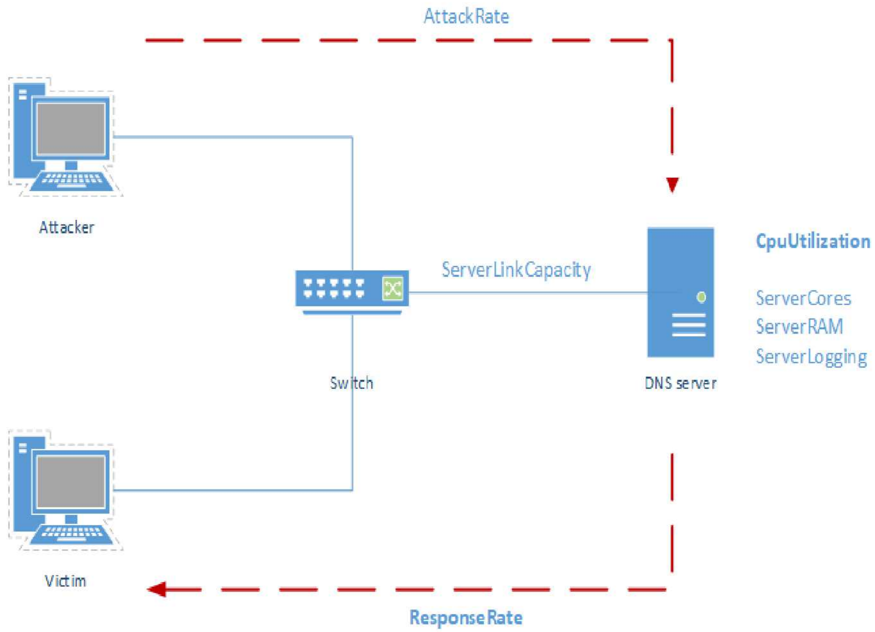


Outputs

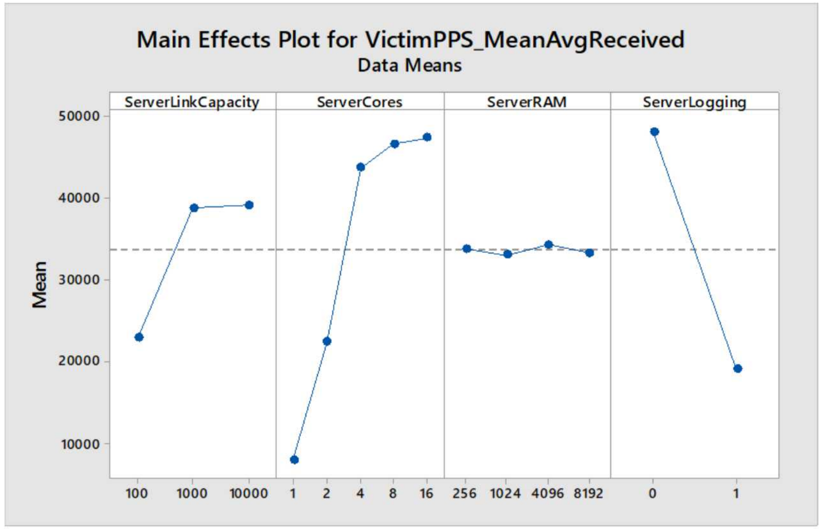
Lots of options…
- Packets
- Host data
- Network data
- Physical Processes

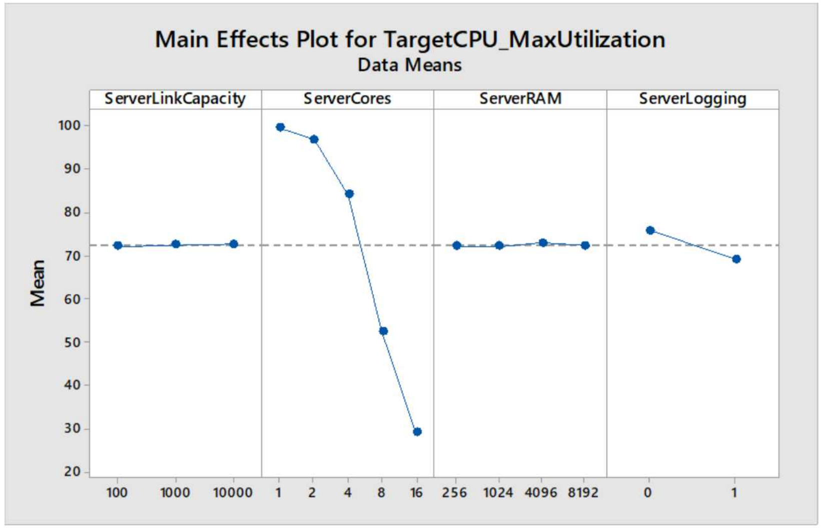
Dakota available at: https://dakota.sandia.gov

Example: Denial of Service Amplification Attack with Full Factorial Parameter Study

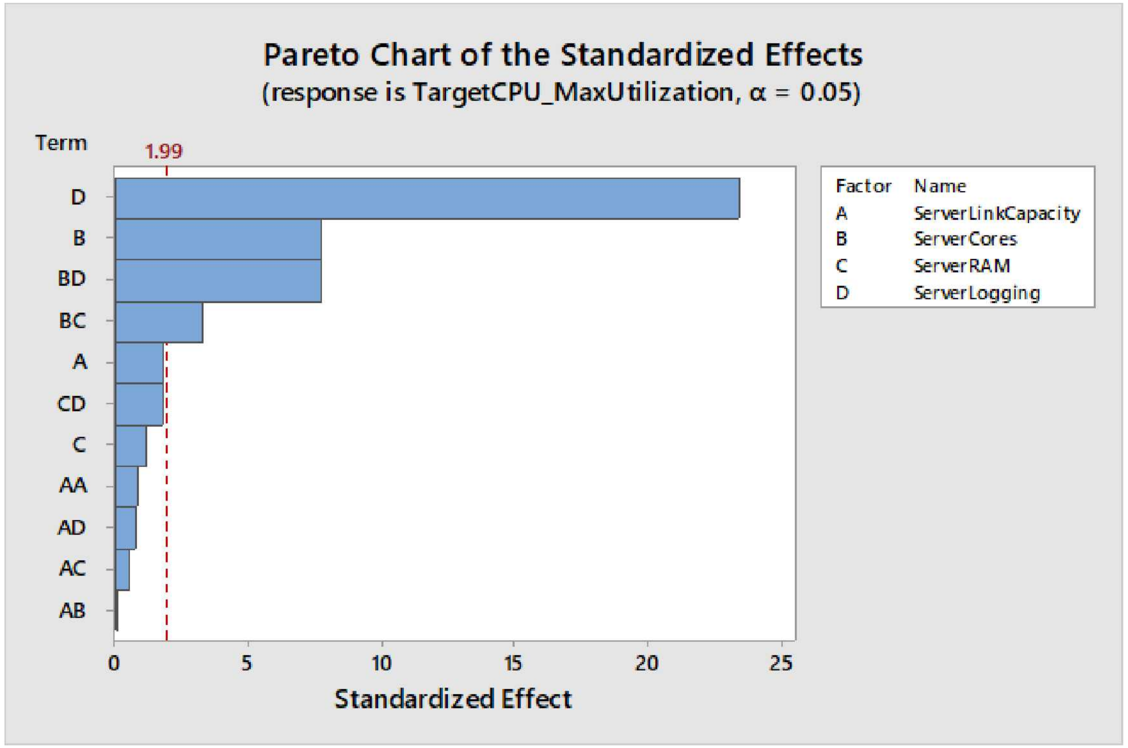


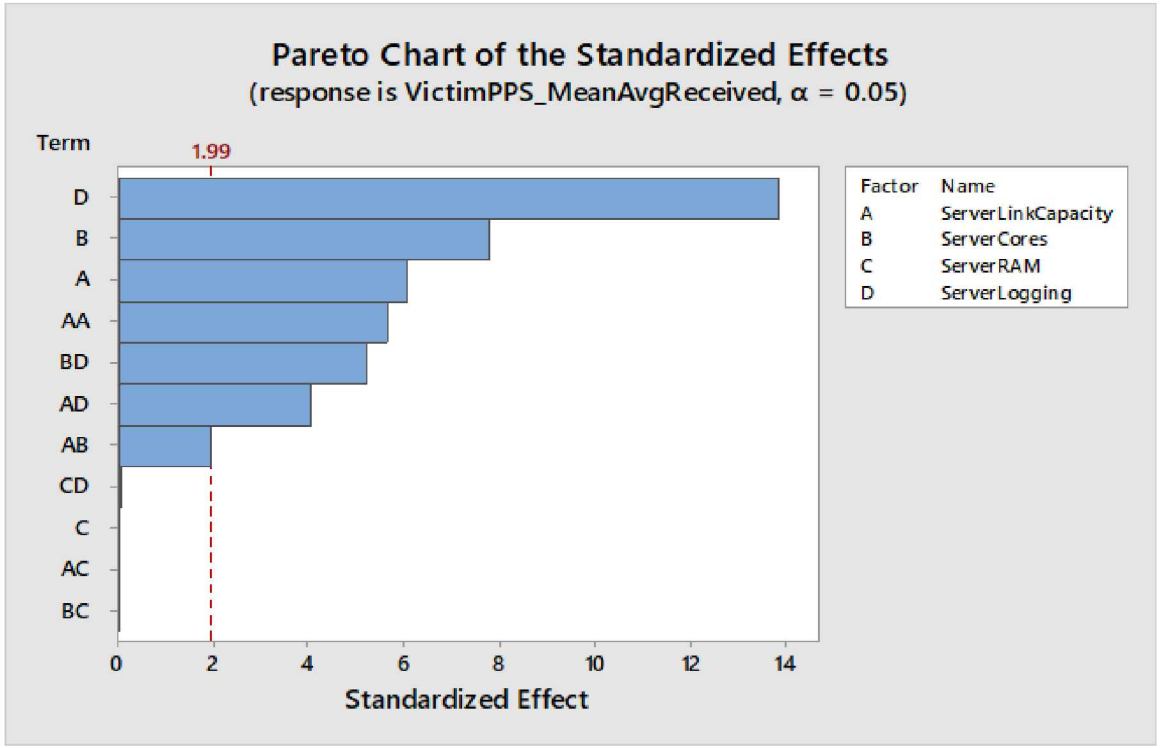
Input parameter	Levels
ServerLinkCapacity	100, 1000, 10000 Mpbs
ServerCores	1,2,4,8,16
ServerRAM	256, 1024, 4096. 8192 MB
ServerLogging	True, False





Example: Denial of Service Amplification Attack with Full Factorial Parameter Study


Pareto Chart of the Standardized Effects
(response is TargetCPU_MaxUtilization, α = 0.05)


Pareto Chart of the Standardized Effects
(response is VictimPPS_MeanAvgReceived, α = 0.05)

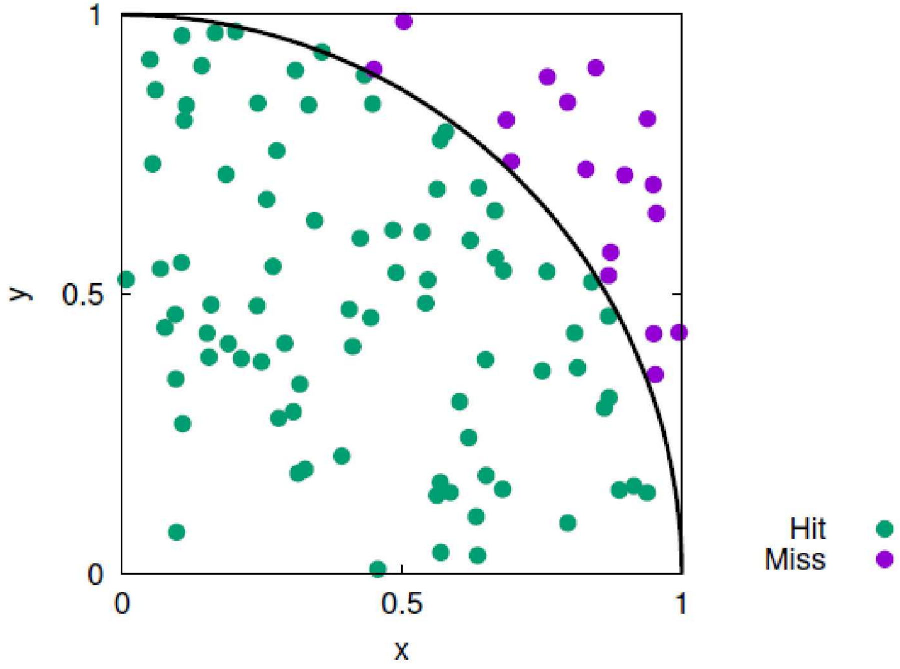
Issues with performing experiments on cyber-emulations

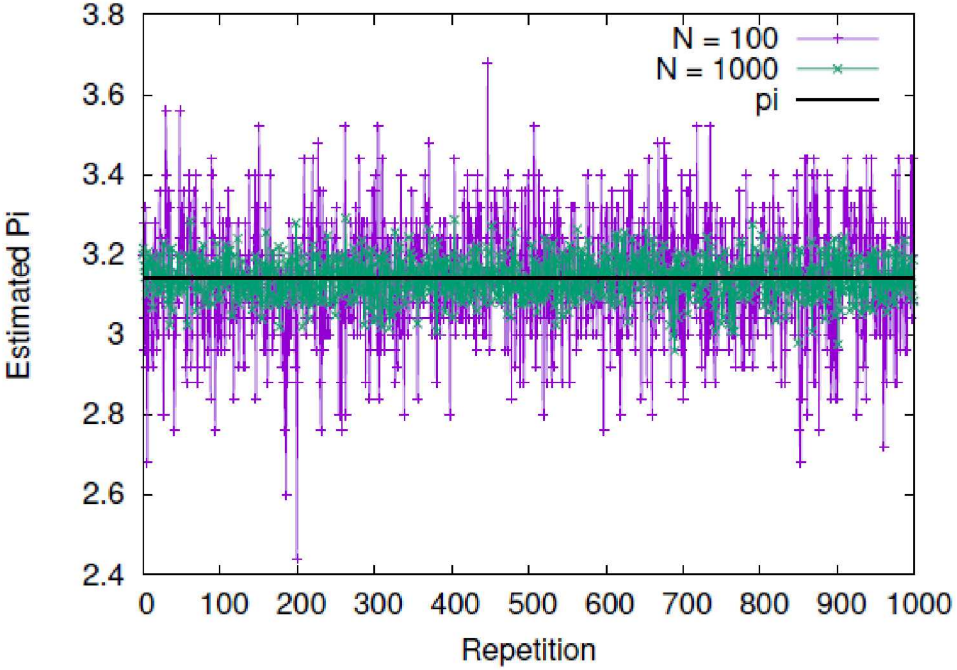
- Emulations may be difficult and time-consuming to set up. For every sample, a number of virtual machines has to be instantiated, initialized, and start communicating with each other.
- Running emulations may also be expensive. If one is emulating a large cyber network with hundreds or thousands of nodes, this will require significant time.
- The UQ community has recently been investigating "multi-fidelity" methods, where many runs (1000s) of a low-fidelity are performed to augment a relatively few (tens) of runs of a high fidelity model.
- In the context of cyber modeling, a network simulator or discrete event model might be the low-fidelity model and the emulation the high fidelity one.
- We are currently investigating the feasibility of multifidelity methods for cyber models.

Multifidelity UQ

Sampling: Most common method to perform uncertainty analysis. Can handle nonlinearities, discontinuities, etc. but has a slow convergence rate.

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



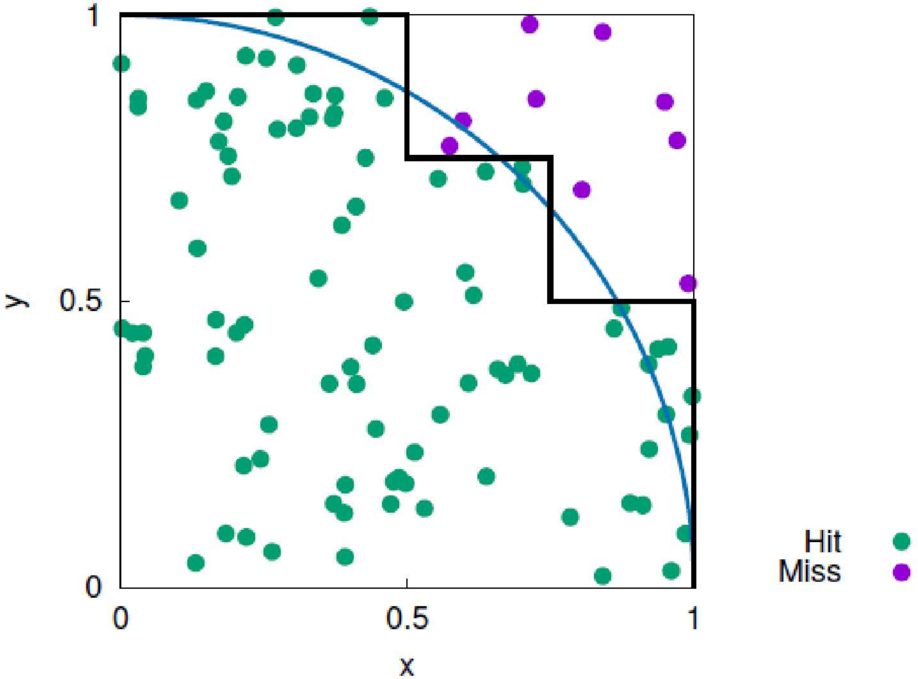


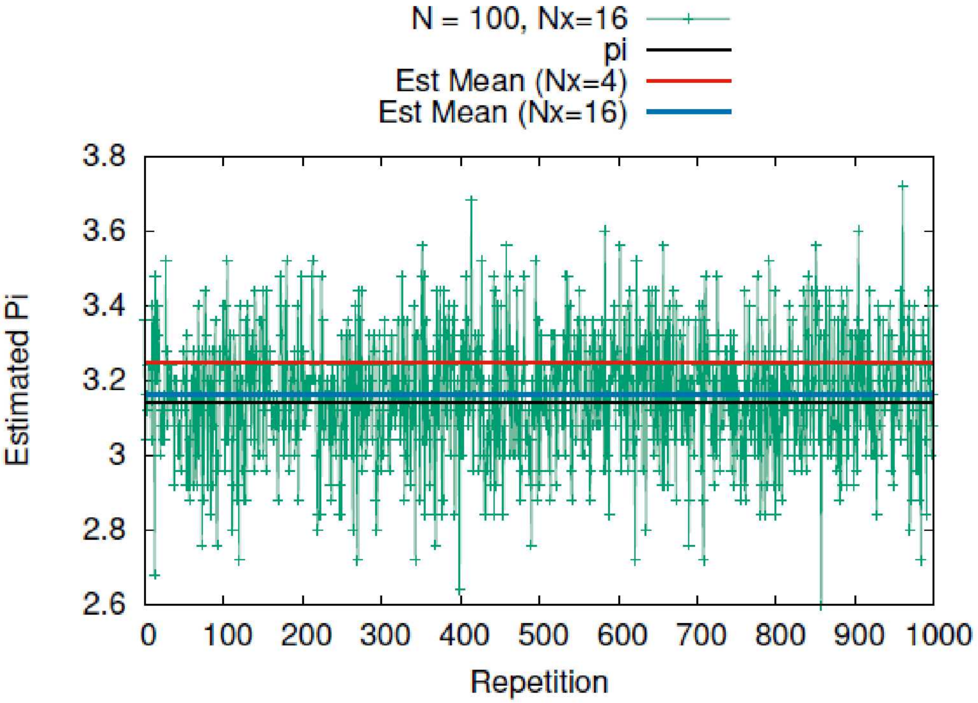
Multifidelity UQ

Whenever a numerical problems **cannot be resolved with infinite accuracy** (discretization error), the MC estimator for a specific Mth 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 $\pi = \dfrac{\#\text{Hit}}{N}$



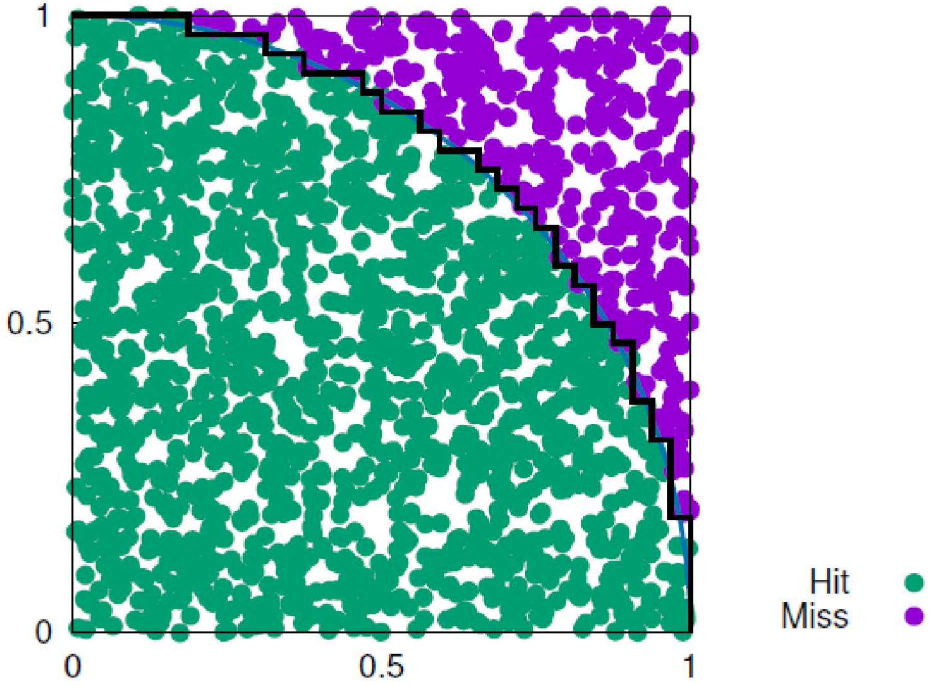


Multifidelity UQ

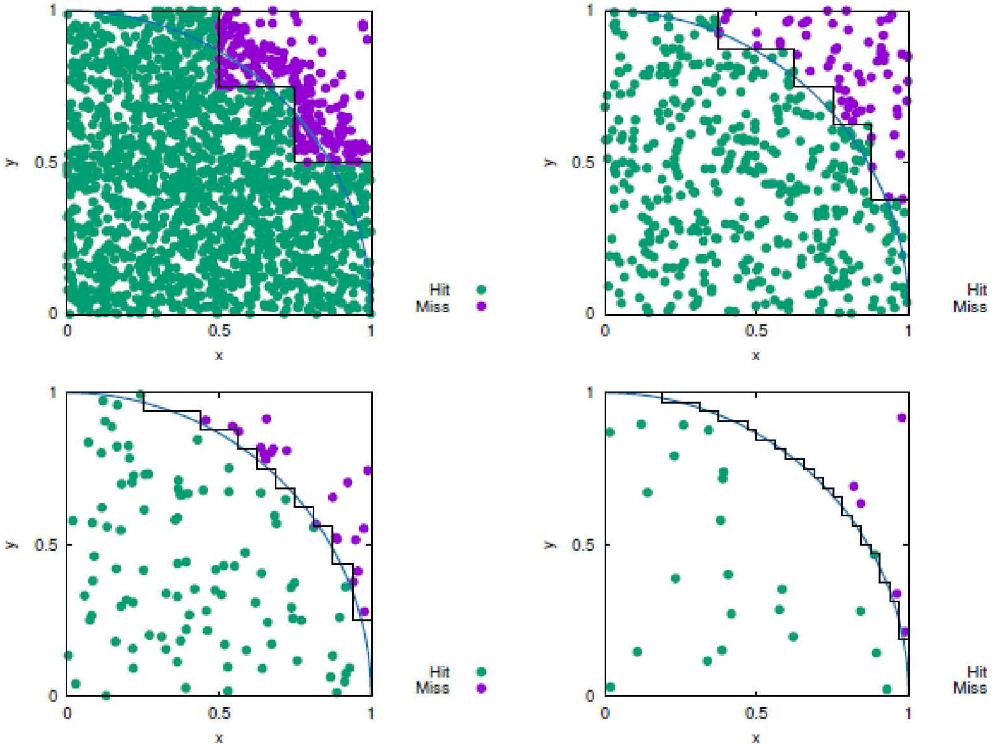
Pivotal idea:

- **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 UQ

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

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

- **Sampling error:** replacing the expected value by a $\underline{\text{(finite) sample average}}$, *i.e.*

$$\mathbb{V}\mathbf{ar}\left(\hat{\mathbf{Q}}^{\mathbf{MC}}_{\mathbf{M,N}}\right) = \frac{\mathbb{V}\mathbf{ar}(\mathbf{Q})}{\mathbf{N}}$$

- **Model fidelity (e.g. discretization):** $\underline{\text{finite accuracy}}$

Accurate estimation ⇒ **Large number** of samples evaluated for the **high fidelity** model

$$\mathbb{E}[Q_M] - \hat{Q}^{MC}_{M,N} \sim \sqrt{\frac{\mathbb{V}\mathbf{ar}(\mathbf{Q_M})}{\mathbf{N}}}\mathcal{N}(0,1)$$

In our network application we operate under the assumptions that

- The emulytics is the highest **unbiased** fidelity model, *i.e.* $(\mathbb{E}\left[\mathbf{Q_M} - \mathbf{Q}\right])^{\mathbf{2}} = 0$
- Our goal is to solely **reduce the variance** of the estimator by introducing low-fidelity evaluations

Multifidelity UQ

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

$$\hat{Q}^{\mathrm{CV}} = \hat{Q} + \sum_{i=1}^{M} \alpha_i \left(\hat{Q}_i - \mu_i \right)$$

- $\hat{Q}_i$ 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: $\mathbf{c} \in \mathbb{R}^{M}$
- $\bar{\mathbf{c}} = \mathbf{c}/\mathbb{V}ar(Q) = [\rho_1 \mathbb{V}ar(Q_1)\,, \ldots, \rho_M \mathbb{V}ar(Q_M)]^{\mathrm{T}}$, where ρ_i is the correlation coefficient (Q, Q_i)

The optimal weights are obtained as $\underline{\alpha}^\star = -\mathbf{C}^{-1}\mathbf{c}$ and the variance of the OCV estimator

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

→ For a single low-fidelity model: $R^2_{OCV-1} = \rho_1^2$

Demonstration of multifidelity approaches to network models: http packet response

Network Configuration

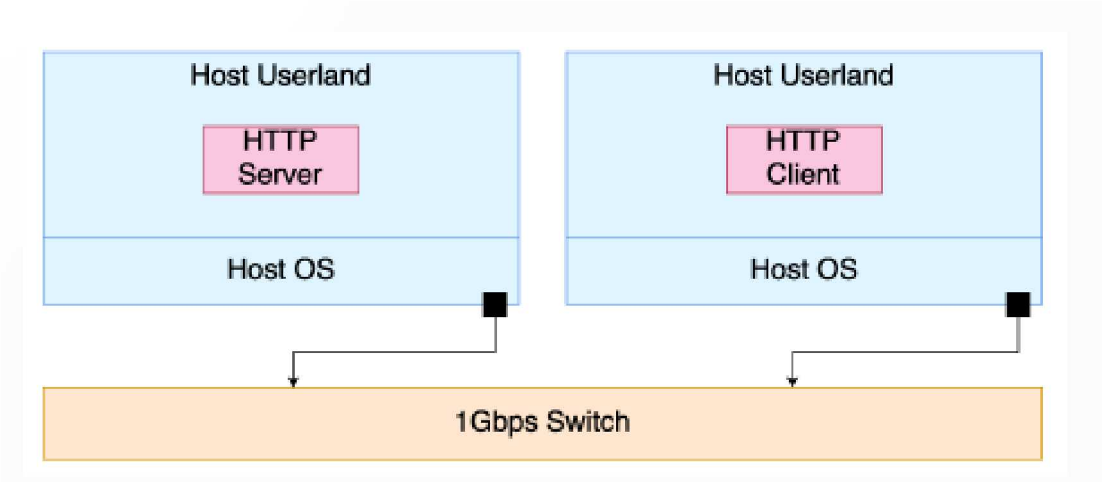
- 1 client - 1 server (possible to extend to multiple clients)
- 100 Requests

Uncertain Parameters

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

Fidelity definition

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



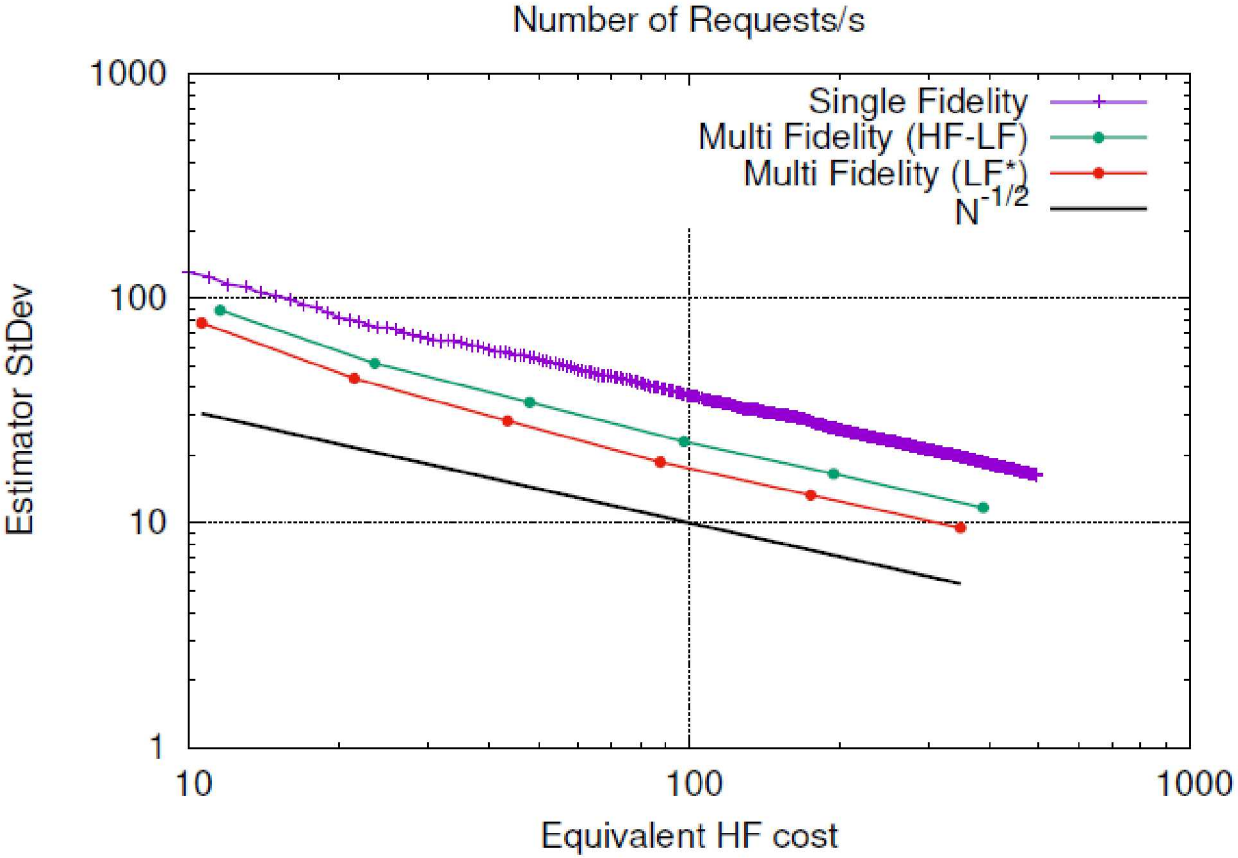
Correlation between models

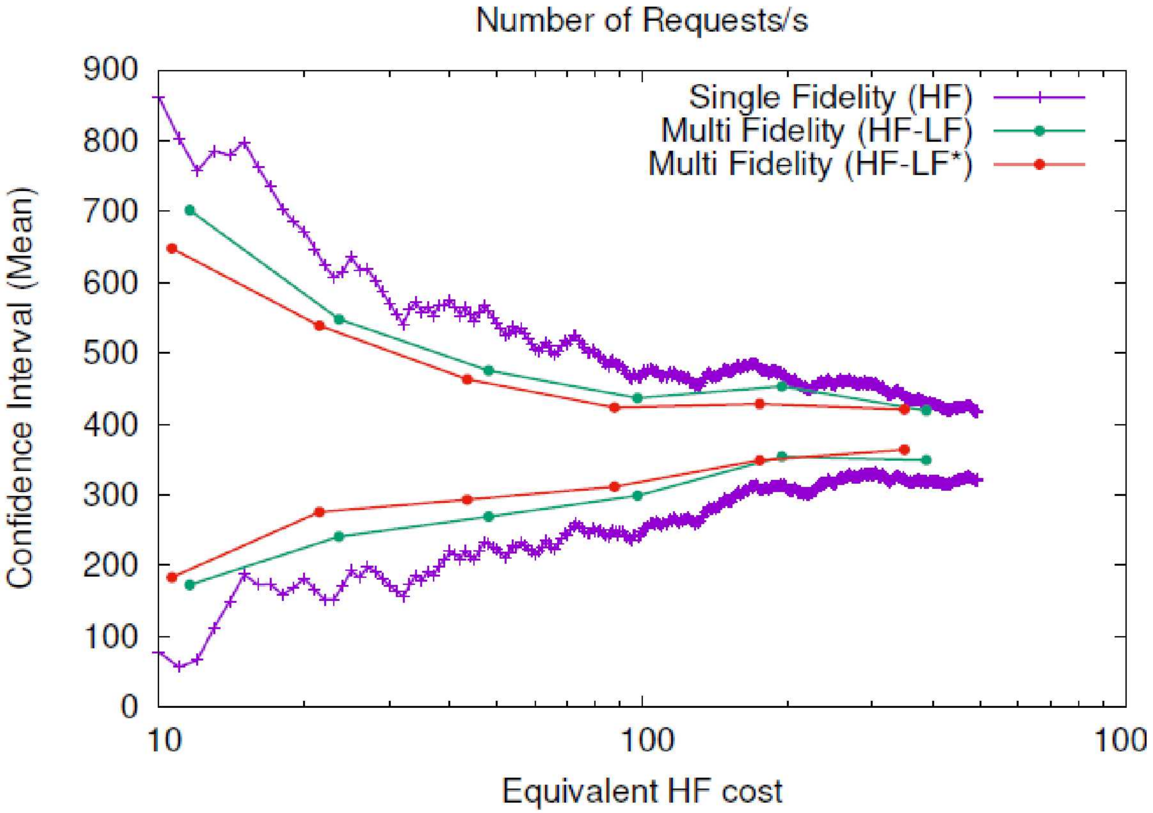
	HF	LF	LF*
HF	1	0.86	0.90
LF	0.86	1	0.99
LF*	0.90	0.99	1

Relative cost between models

Model	Cost
HF	1
LF	0.016
LF*	0.002

Multifidelity Results





Graphs show 4-fold reduction in variance when estimating the ***mean number of http requests completed per second*** for the same computational budget

Conclusions

- Computational workflows around Emulytics models are starting to be developed similar to what the computational simulation community has done.
- Uncertainty quantification and experimental design are important components in understanding the sensitivities of a model and identifying the range and distribution of model output.
- There are several issues performing UQ for emulations, including replicability and large variation in performance across runs; difficulty of setting up multiple emulations concurrently; and large, discrete parameter spaces.
- The use of new approaches to UQ such as multifidelity methods is interesting and worthy of further investigation.