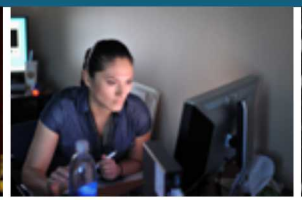
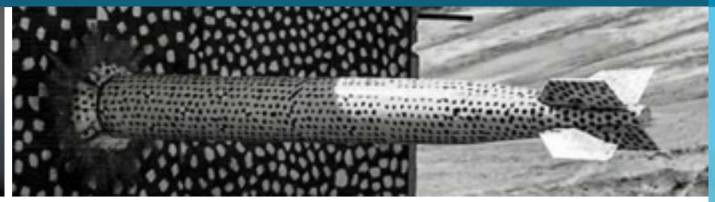
SAND2020-6718C

Sandia National Laboratories

Information Theoretic Metric of Verification System Performance





AUTHORS

Jason Reinhardt, Michael Hamel, Ben Bonin, Eva Uribe

PRESENTED BY

Jason Reinhardt

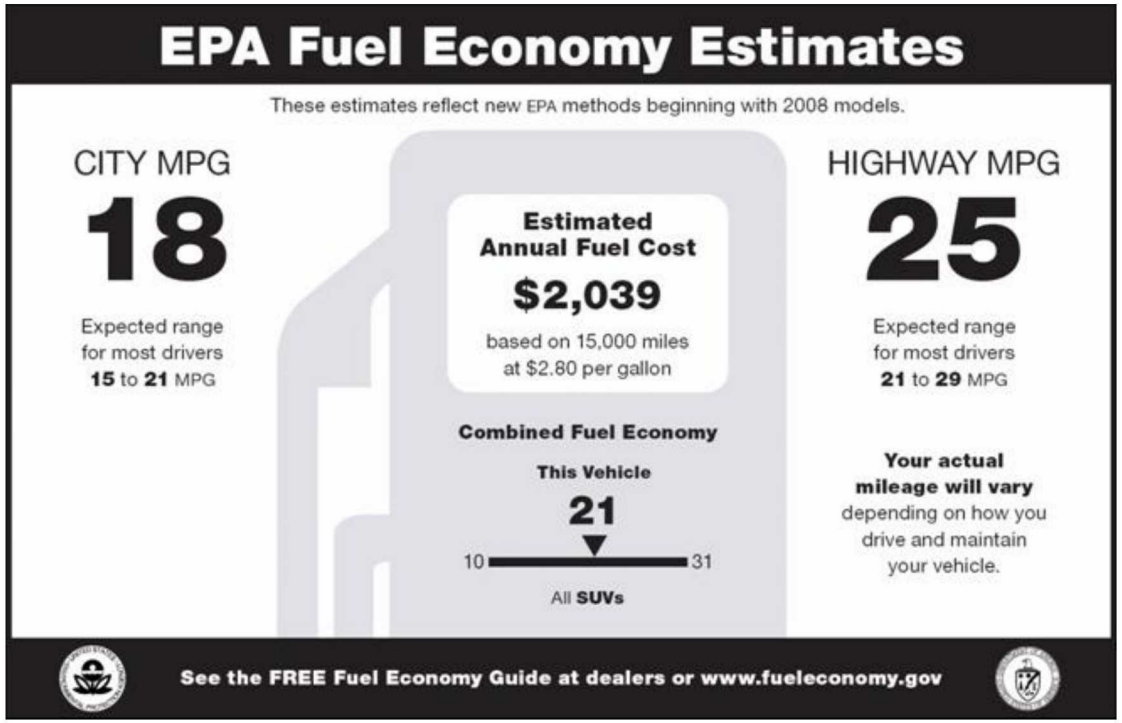
U.S. DEPARTMENT OF ENERGY NNSA

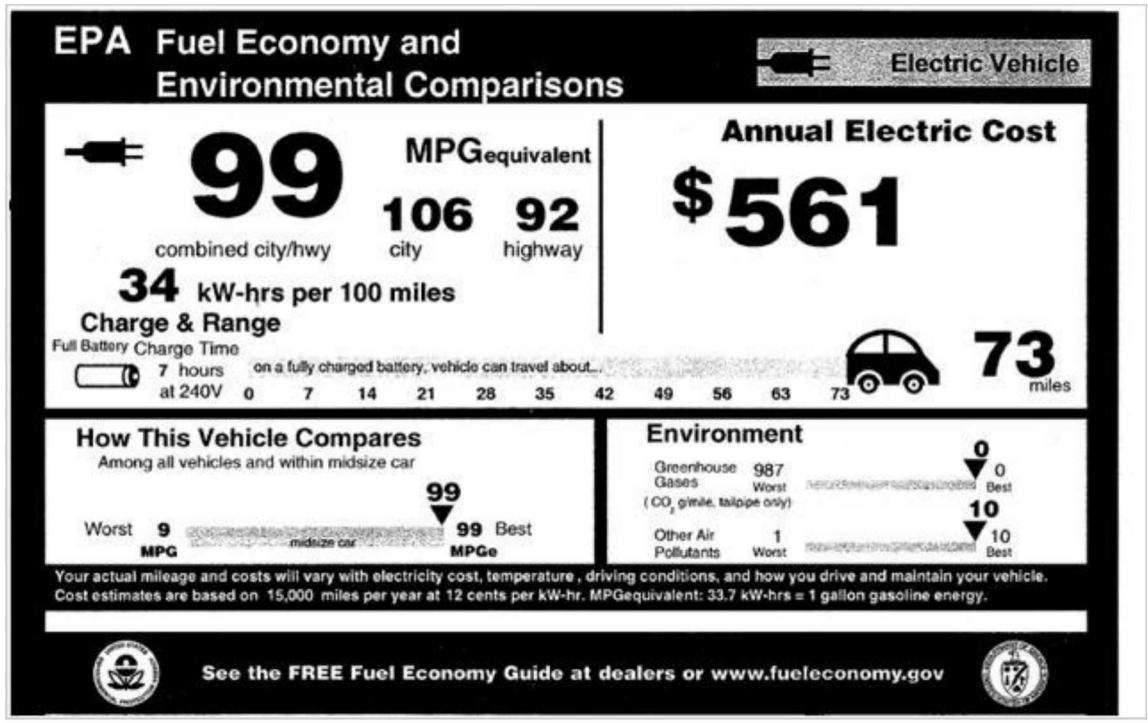
Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Partnerships for Developing Arms Control Verification Metrics

Technology and protocol agnostic metrics promote common understanding of performance and principles, and can provide new opportunities for cooperative efforts to meet verification challenges.

Developing such metrics and assessment mechanisms in collaboration with potential or actual treaty and agreement partners and stakeholders may provide a basis for collaborative advancement.





Information Barriers and the Core Dilemma

The requirements for an information barrier are two-fold:

- Protect the host, or inspected party, by guaranteeing that sensitive data cannot be transmitted to any other party through a measurement.
- Provide the monitoring party that the measurement validates a claim by the host, using authentic and accurate data.

Information communication through an observed verification signal (Y) is indicated by an updated monitor's probability distribution over the protected quantity (X).

The desired state for no information communication through the observed signal is:

$$p(X|Y) = p(X)$$

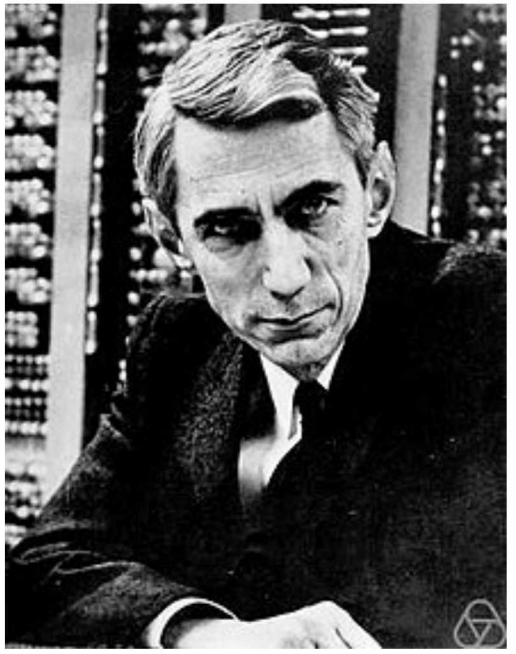
The challenge of developing such systems is a probabilistic (in)dependence issue: ***How do we create a verification signal (Y) that minimizes probabilistically dependence on protected quantity (X)?***

Information theoretic approaches may provide verification technology and protocol agnostic methods for assessing the potential information transmission in a verification measurement/signal.

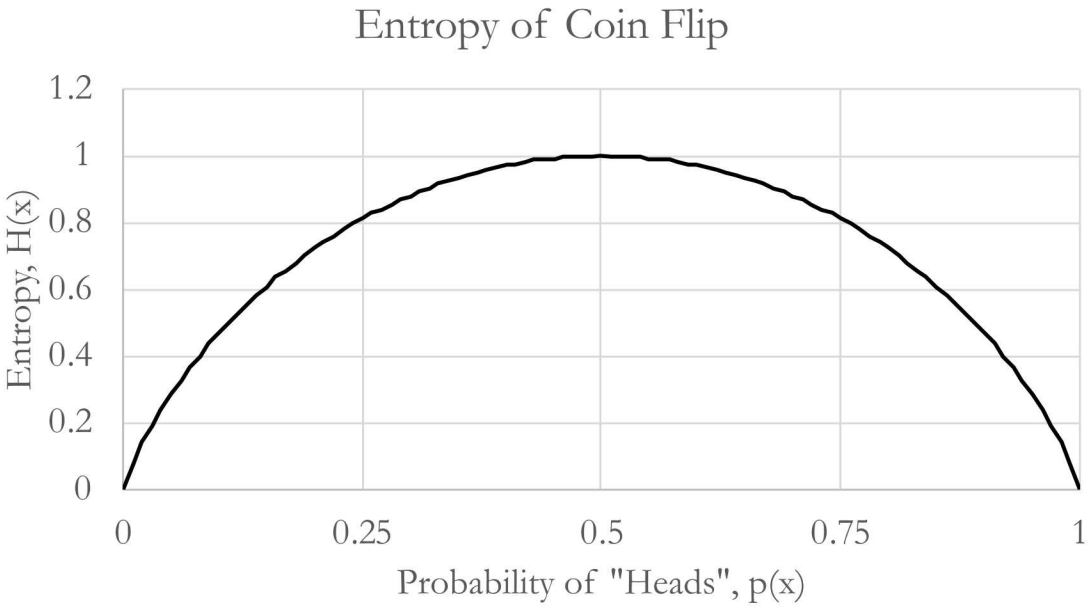
Information Theory Fundamentals

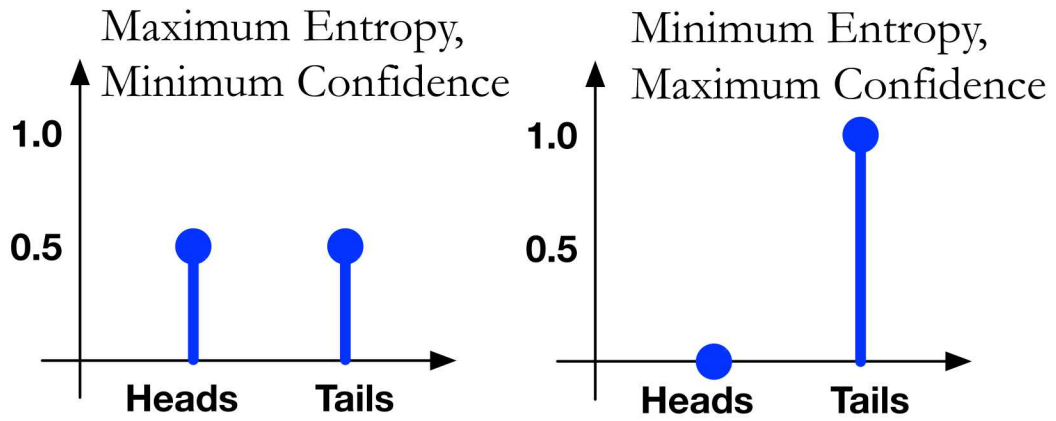
Information theory provides a mathematical basis for understanding the transmission and quantification of information.

- Originally focused on limits of transmission channels and storage
- Defines the concept of "information entropy" as a measure of disorder in a probability distribution
- Number of 'bits' required, on average, to communicate a sequence of observations from a random process



Claude Shannon





$$H(p_X) = -\sum_{x \in X} p_X(x) \log_2 (p_X(x))$$

Image: Jacobs, Konrad, https://en.wikipedia.org/wiki/Claude_Shannon#/media/File:ClaudeShannon_MFO3807.jpg

Deriving a Measure of Information Barrier Performance

Information shared between two uncertain quantities can be expressed as the ***mutual information.***

$$I(X;Y) = \sum_{x \in X, y \in Y} p(x,y) \log_2 \left(\frac{p(x,y)}{p(x)p(y)} \right)$$

The ***Kullback-Liebler divergence*** can be used to show the information gain of making an observation.

$$D_{KL}\big(p(X|y)||p(X)\big) = \sum_{x \in X} p(x|y) \log_2 \left(\frac{p(x|y)}{p(x)} \right)$$

Mutual information can be expressed in terms of the Kullback-Liebler divergence.

$$I(X;Y) = \sum_{y \in Y} p(y) D_{KL}\big(p(X|y)||p(X)\big)$$

This is the ***expected*** Kullback-Liebler Divergence, and is a measure of how much information is expected to be gained about a protected quantity (X) from the monitors activities to verify the host's claims through observing (Y).

Interpreting the Measure in Context

Let X be a protected quantity whose possible outcomes that are sensitive and must be protected (e.g., inspected item design details), and let Y be a signal that an monitor will collect (e.g., from detection equipment specified in protocol).

Derived from design parameters for detection systems

$$I(X;Y) = \sum_{y \in Y} p(y) D_{KL}\left(p(X|y)||p(X)\right) \rightarrow 0$$

Monitor's beliefs over inspected item information

Ideal Condition

The measure $I^{\star}(X;Y)$ can be the basis for information barrier performance assessments and methods:

- Metric is measured in 'bits'
- Design parameters can be assessed empirically via testing and evaluation of verification systems and protocols
- Ideal can only be achieved when X and Y are probabilistically independent
- ***Depends on the beliefs of the monitor, which the monitor is unlikely to know ahead of time***

Proposed Information Barriers System Performance Metric

We can now define a protocol and technology independent, system level, performance metric for information barriers and verification systems: the maximum mutual information or expected Kullback-Liebler divergence.

The value of the metrics can be found by measuring the classification performance of the system ($p(Y|X)$) and solving the mathematical program shown.

This finds the worst case difference between information leakage and verification confidence, over all possible monitor beliefs of X.

It may be necessary to define a standard set of inspected item properties that must be protected (X), as a reference set for estimating the metric collaboratively.

$$I^{\star}(X;Y) = \max_{f \in \Delta_X} \sum_{y \in Y} p(y) D_{KL}\big(p(X|y)||p(f)\big)$$

$$s.t.$$

$$\sum_{x} f_X(x) = 1$$

$$\sum_{x \in X} p(x|y) \log_2 \left(\frac{p(x|y)}{p(x)}\right) = D_{KL}\big(p(X|y)||p(f)\big)$$

$$\frac{p(y|x)}{p(y)} f_X(x) = p(x|y) \quad \forall x, y \in X, Y$$

$$\sum_{x \in X} p(y|x) f_X(x) = p(y) \quad \forall y \in Y$$

$$0 \leq f_X(x) \leq 1 \quad \forall x \in X$$

An Example: Setup

A hypothetical verification regime requires a host to present a sealed container with a treaty accountable object (TAI) inside.

A legitimate treaty accountable object consists of a mixture two isotopes: Cs-137 and {Co-60 or Y-88}

The monitor measures the γ-ray radiation energy spectra to confirm that the container holds a legitimate object.

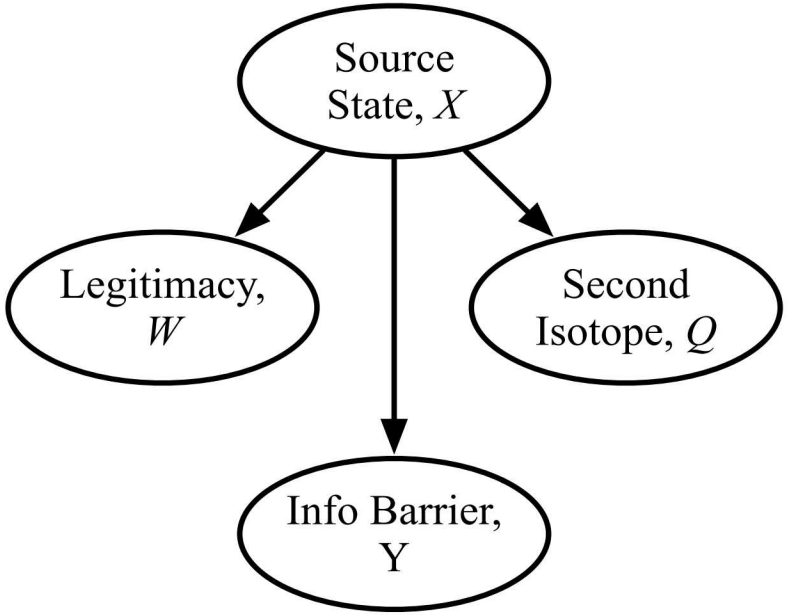
The host would like to hide exactly which treaty accountable object is being presented, but convince the monitor that is legitimate.

Source State	Legitimate or Spoof Item?	Activity of Cs-137 (μCi)	Activity of Co-60 (μCi)	Activity of Y-88 (μCi)	Mean CPS in Window
0	Empty (Background)	0	0	0	4.97
1	Legitimate	7.263	9.3	0	1.51E3
2	Legitimate	7.263	0	10.0	1.45E3
3	Spoof	7.263	0	0	1.02E3
4	Spoof	7.263	14.7	0	1.75E3
5	Spoof	7.263	0	12.8	1.70E3

An information barrier is proposed that only reports if the photon counts in a certain energy range of γ-ray energies is within a specified set of values (n standard deviations of the mean).

How do we choose the parameters of the information barrier (e.g.. value of n)?

An Example: Simulations and Models

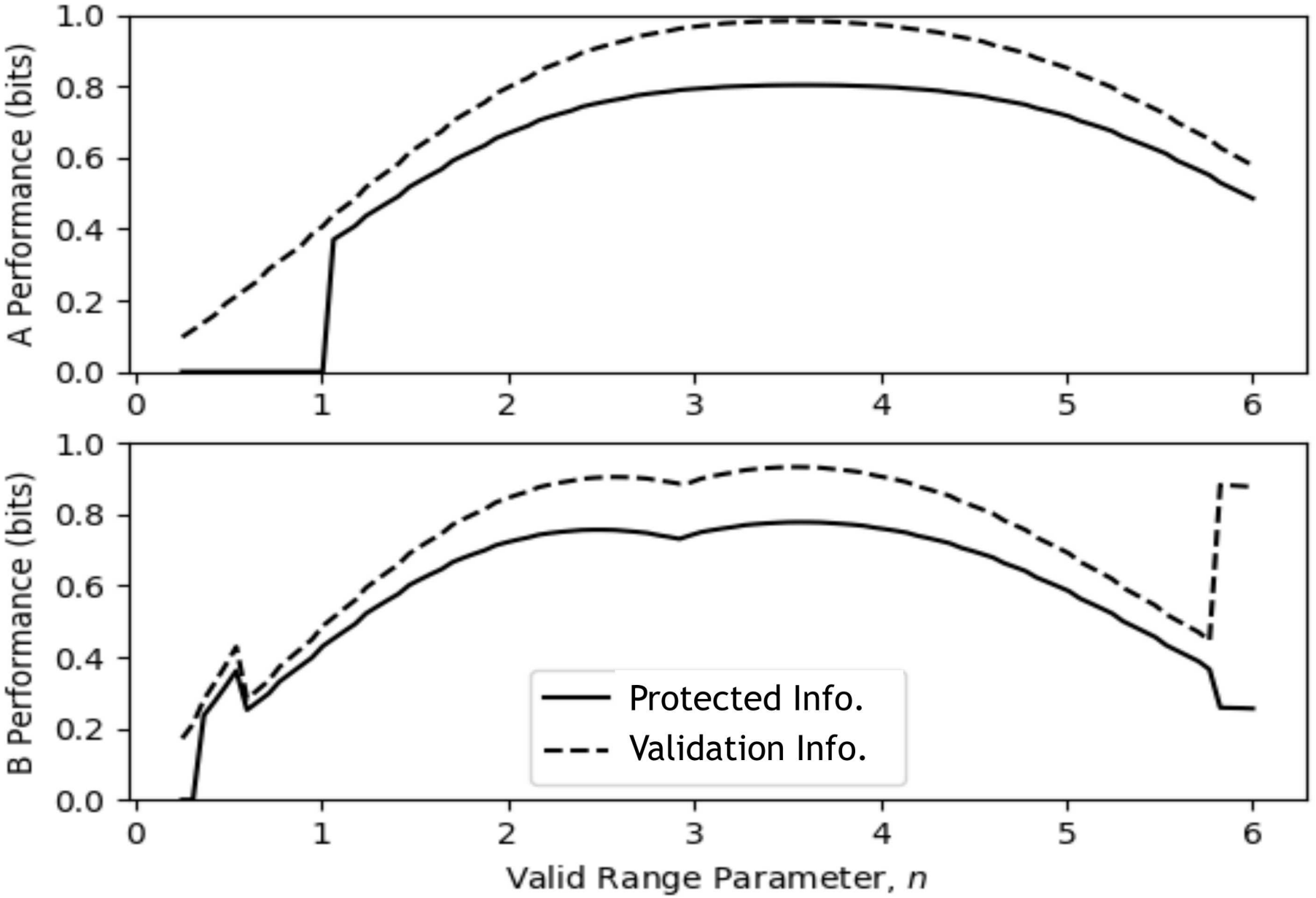


Spectra for each of the possible source states objects were simulated using GADRAS

A simple Bayesian model was constructed in Python, using notional data

Optimizations were run using the SciKit toolbox to measure the value of the proposed metric for various values of n

An Example: Results – Information Transmitted per Design Parameter, n



Preliminary Conclusions for System Metrics and Approaches

Information Theoretic approaches provide a principled framework for verification system analysis that may yield significant insights into protocol and technology design.

- Built on first principles from probabilistic analysis methods
- Focuses on state of knowledge and confidence and optimizing the inherent trade-offs
- Allows for explicit treatments of uncertainty and dependency
- Serves as a basis for game-theoretic and analytic approaches
- Rely directly on laboratory measurements of system performance

As an example, the mutual information, or expected Kullback-Liebler divergence, is proposed as a method for assessing and understanding information barriers challenges and performance from first mathematical principles.

Many verification problem areas and applications remain to develop consistent, high-level approaches to understand performance, design limits, trade-offs, etc. Results may inform:

- Understanding what's possible, and identifying theoretical limits
- Next step R&D (analyze current approaches, standardization of test problems/scenarios)
- Protocol design and verification technology selection