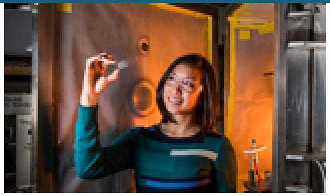
SAND2021-3306PE

Anomaly Detection and Surety for Safeguards Data



PRESENTED BY

Natacha Peter-Stein[1], David Farley[1], Constantin Brif[1], Nicholas Pattengale[1], Chase Zimmerman[1], Yifeng Gao[2], Jessica Lin[2], Mitchell Negus[3], Rachel Slaybaugh[3]

[1]Sandia National Laboratories, Albuquerque, NM and Livermore, CA
[2]George Mason University, Fairfax, VA
[3]University of California, Berkeley, CA

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.
XXXX

Introduction

Nuclear Safeguards – Data-rich Field

- Ideal for the application of modern data analytics techniques
- Technologies necessary for IAEA implementation not sufficiently mature

Data Analytics Project

- Multidisciplinary teams at ORNL, LANL, and SNL working together to advance the suite of data analytic capabilities to support safeguards activities at declared facilities
  - Data conditioning
  - Safeguards questions development
  - Red teaming exercises

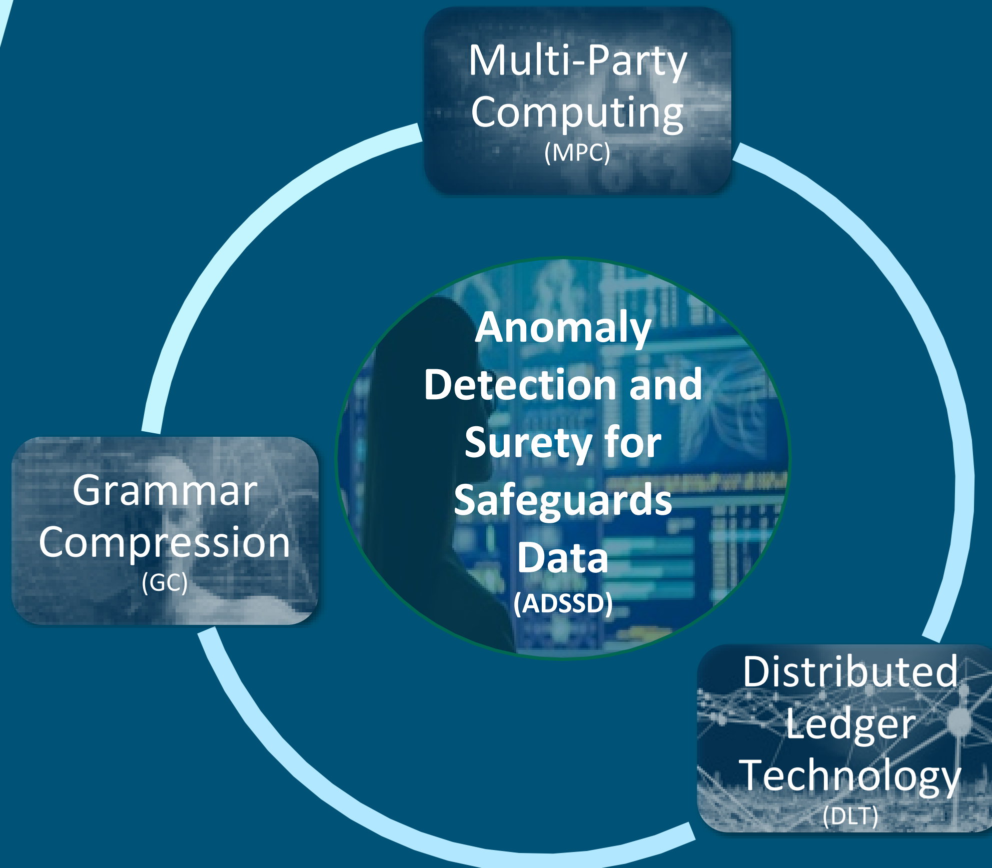
The SNL team is focused on data surety and anomaly detection

"to ensure Continuity of Knowledge and improve timely diversion detection"

Project Overview

Goals

Technical Approach



Investigation of three core data analysis and management methods and their applicability for international safeguards

- ***A****nomaly detection based on the GC method*
- ***D****evelop and test a novel safeguards data authentication, integration, and analysis workflow on the foundation of DLT*
- ***H****ow operator data could assist in drawing safeguards conclusions in a MPC environment*

Project Deliverables

Title	Description	Status
Use Case documentation	Report on proposed safeguards use cases	Complete (9/15/2019)
Prioritized anomaly detection methods	Report on the prioritization method and selected anomaly detection methods	Complete (6/30/2019)
Down-selection of technologies and data for prototype DLT system	Report on selected type of prototype DLT system	Complete (6/30/2019)
MPC Viability Assessment	Report on test scenarios with known anomalies to evaluate how easily anomalies in raw data sequences convert through a garbled circuit	Complete (6/30/2019)
Implement anomaly detection methods	Software tool implementing selected anomaly detection methods	Complete (9/30/2020)
First prototype DLT system	Software tool implementing first version of prototype DLT system	Complete (9/30/2020)
Application of MPC-based protection to actual data	Report on application of MPC approach to actual data streams (e.g., MINOS)	Complete (9/30/2020)
Demonstration of the full system	Software tool(s) implementing GC anomaly detection, MPC-based data protection, and DLT-based data surety that works with the integrated system and with common data streams	9/15/2021

Why Grammar Compression Based Anomaly Detection is Useful for Safeguards Data

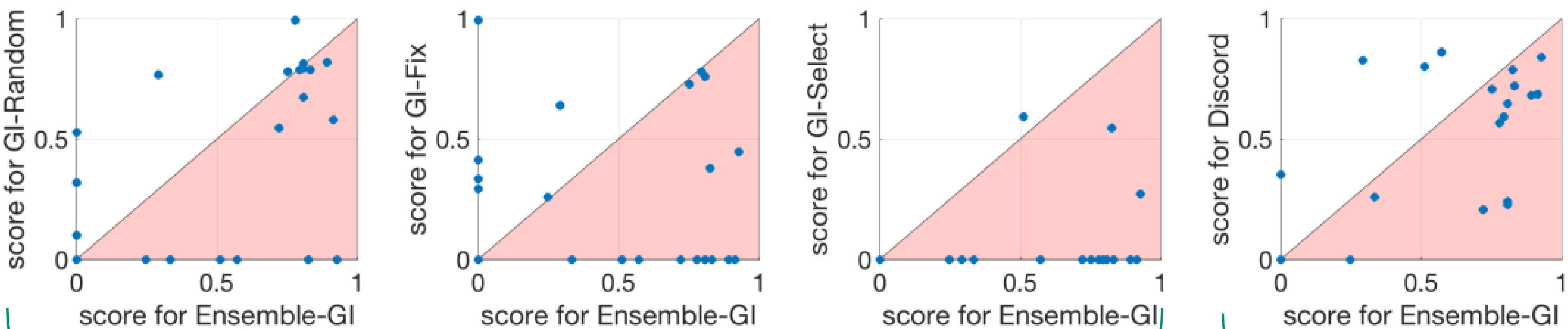
- We are developing a practical method for effective and efficient detection of anomalies in multivariate time-series data obtained from safeguards used for monitoring of civilian fuel cycle activities.
- The key component of the proposed approach is the cutting-edge method of unsupervised anomaly detection based on ***Grammar Compression*** (GC).
- This method has a number of crucial advantages important for analysis of safeguards data.

Challenges Posed by Safeguards Data	Capabilities of GC
Safeguards generate large amounts of data (about one million reports generated each year need to be analyzed).	GC is a cutting-edge technique that ***scales linearly*** with data size and has demonstrated superior performance for a number of real-world applications.
We need to address the multivariate character of data obtained from heterogeneous sensors, including video cameras, radiation detectors, electronic seals, etc.	GC can be extended to include the capability for detection of correlated (sub-dimensional) anomalies in high-dimensional data.
Data analysis involves imprecisions (approximation errors) associated with the extraction of discrete features from continuous waveforms.	GC approximates time-series data in a way that lower -bounds the true distance for the original time-series. Moreover, GC can be extended to incorporate ***Ensemble Learning*** for improved robustness against approximation errors.
Training datasets with labeled “normal” and “abnormal” events are lacking.	GC employs unsupervised learning, i.e., compares the data against themselves, and therefore does not require a labeled training set.

Advances: Ensemble Grammar Compression

- We combined GC with ***Ensemble Learning*** to achieve robust and efficient anomaly detection.
- Ensemble Learning uses averaging over multiple algorithm executions with randomly selected values of discretization parameters. This achieves detection accuracy comparable to that of exact algorithms while maintaining a linear time complexity. Paper presented in EDBT 2020 (March 2020).

To evaluate performance of ensemble GC we used 6 different datasets and 25 time series for each type of data. Plots below show comparison against four baseline methods for one of the datasets. A point in the lower triangle corresponds to a superior performance by ensemble GC compared to the baseline method.



Compared against three variations of parameter value selection approach (random, fixed, and optimized) in the standard GC method

Compared against *Discord Discovery*, the state-of-the-art method that scales quadratically with data size

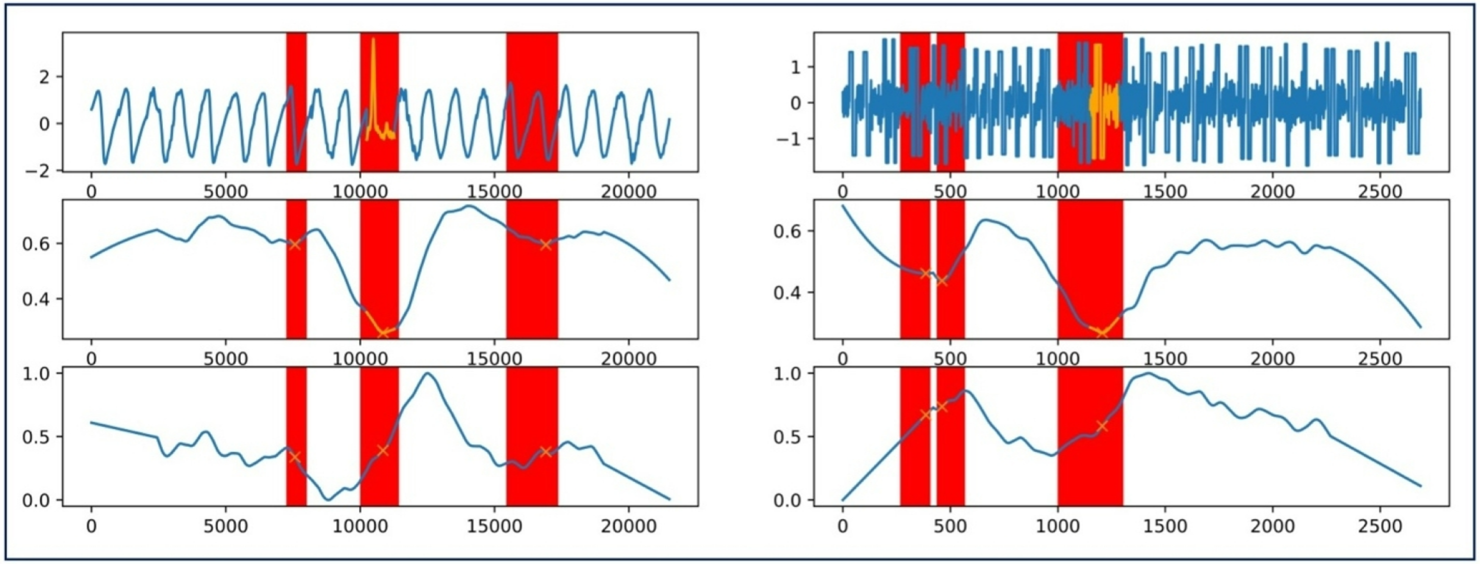
Performance comparison: Score averaged over 25 time series

Ensemble GC	GC-Random	GC-Fix	GC-Select	Discord
0.473	0.372	0.241	0.056	0.400

Advances: Motif-Based Anomaly Detection

Motivation: Detecting Anomalies on Extra-Long Scale

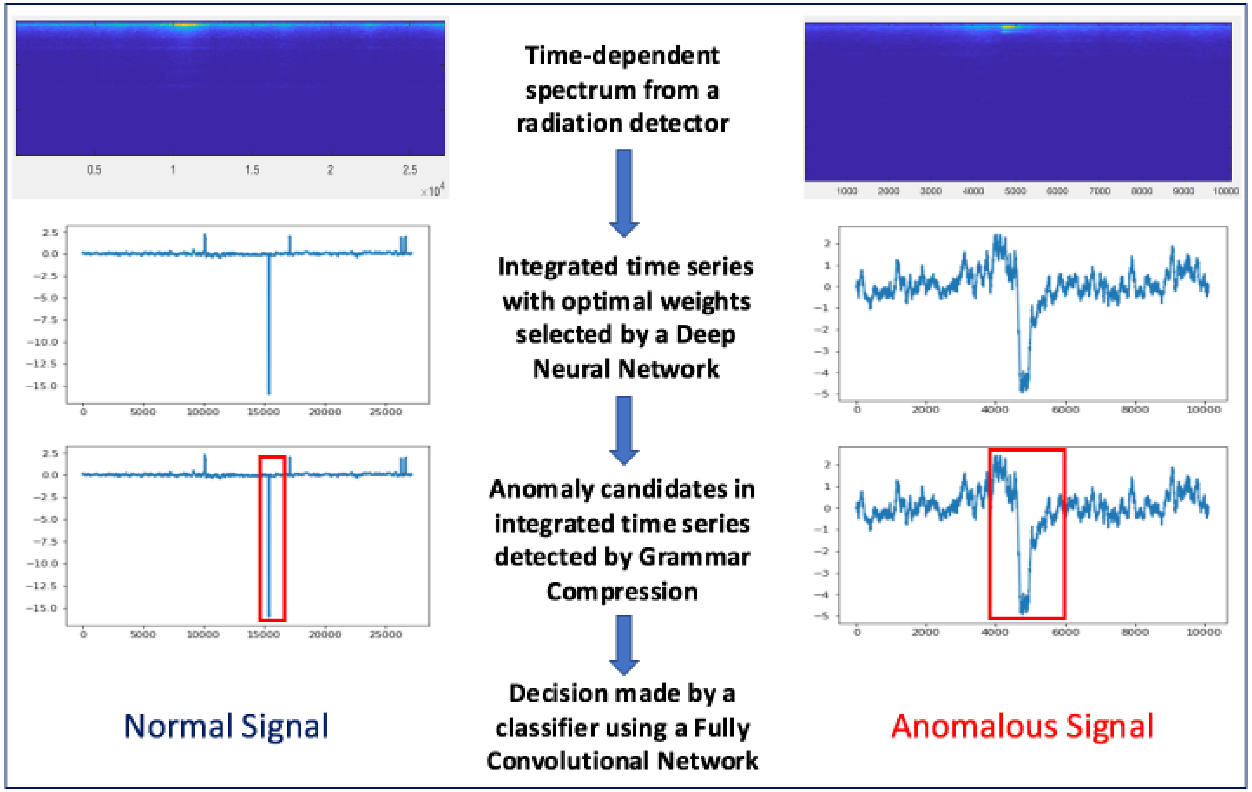
- GC is a “greedy” algorithm that focuses on variations that occur on a short time scale.
- To detect anomalies on extra-long scale (time series with millions of data points) we leveraged a new variable-length motif discovery algorithm, Hierarchy-based Motif Enumeration (HIME).
- Motifs are recurrent patterns in a time series.
- Motif discovery can be used as a key step in anomaly detection — subsequences that contain least number of frequent motifs are anomaly candidates.
- Specifically, the new method computes a motif correlation density curve (MCDC) whose minima indicate anomaly candidates. The length of each anomaly candidate is evaluated by computing the derivative of the MCDC around a minimum point.



Advances: Anomaly Detection in Multivariate Data

Motivation: Detecting Anomalies in Data from Radiation Detectors

- A new method is designed to work with time-dependent spectral data such as those obtained from radiation detectors. Specifically, a radiation detector records data at multiple spectral components (gamma ray energies), with the number of counts recorded for each energy being one of the multiple variables.
- This new method combines deep learning (DL) and grammar compression (GC).



Advances: Anomaly Detection in Multivariate Data

Application to Detection and Identification of Radioactive Materials

- For training and testing of the new method, we utilized a simulated radiation detection dataset (radDetect) developed by ORNL for open use.
- Results of a numerical experiment with radDetect data:

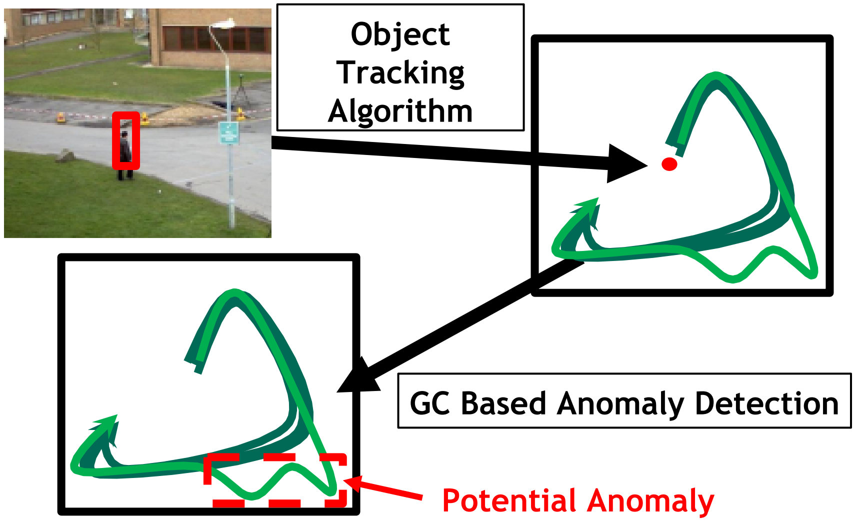
Radioactive material		Correct prediction %
No material (normal data)	4900	100%
Highly enriched uranium (HEU)	800	87%
Weapon grade plutonium (WGPu)	800	82%
^{131}I	800	80%
^{60}Co	800	92%
^{99m}Tc	800	80%

Actual Prediction	normal	HEU	WGPu	^{131}I	^{60}Co	^{99m}Tc
normal	1065	1	0	1	0	1
HEU	0	128	0	2	1	1
WGPu	0	4	149	9	4	11
^{131}I	0	7	17	142	2	15
^{60}Co	0	1	2	2	174	5
^{99m}Tc	0	6	14	21	8	129

Future Plans for GC Extensions and Tests

Using GC to Detect Anomalies in Video Data

- A straightforward approach is to consider each pixel as a separate time series.
- The proposed approach is to use tracking of moving objects: first, use an object tracking algorithm to extract trajectories of all moving objects. Second, use GC to detect anomalous trajectories.



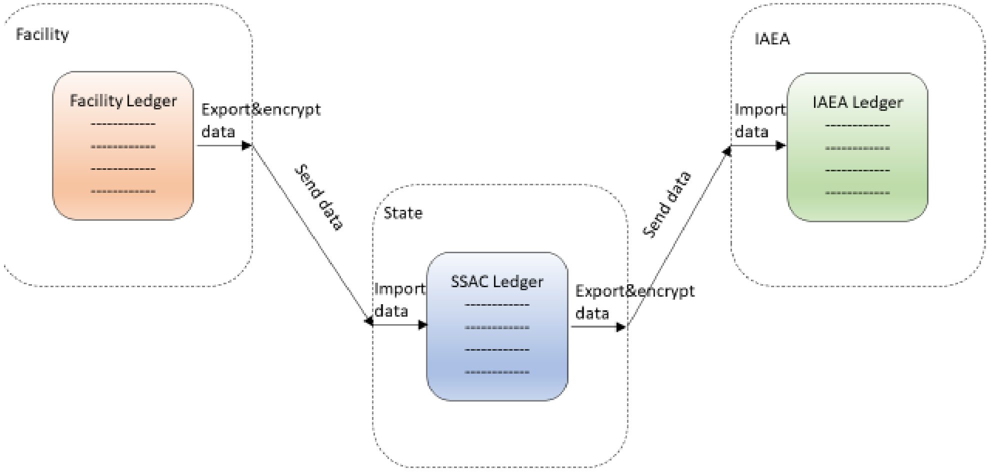
Testing GC extensions on MINOS data

- Some of the MINOS datasets are of particular interest to us in order to test & evaluate the developed GC-based anomaly detection methods:
  - ORNL Distributed Fiber Optic Sensor (DFOAS),
  - ORNL MUSE,
  - ORNL Ground Truth

Conduct "Blue Team/Red Team" exercises in Year 3 (FY21)

- We have provided the Ensemble GC software package to LANL for testing, and expect to collaborate with them on analysis of their results.

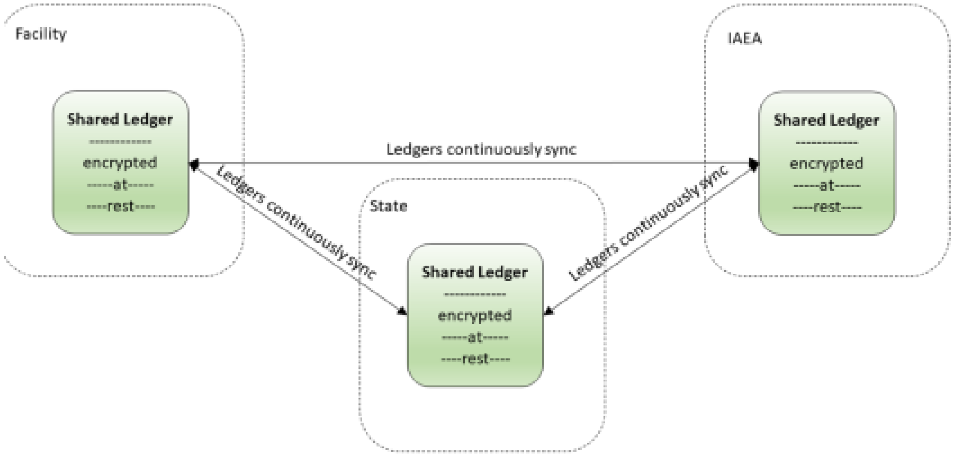
DLT Concept Recap



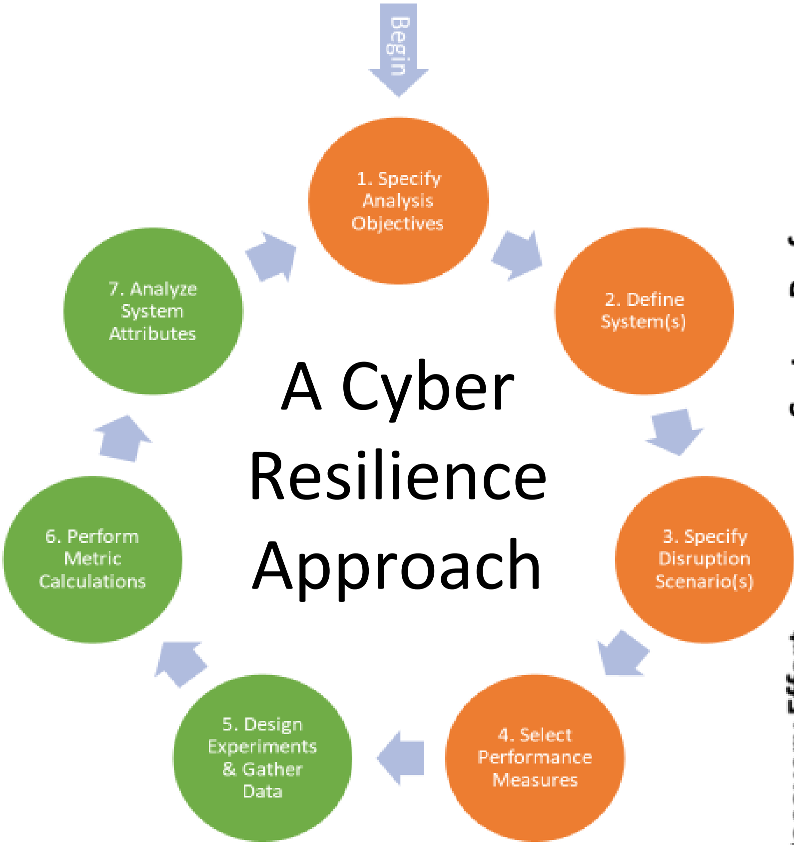
We consider adoption tiers, with varying levels of potential impact

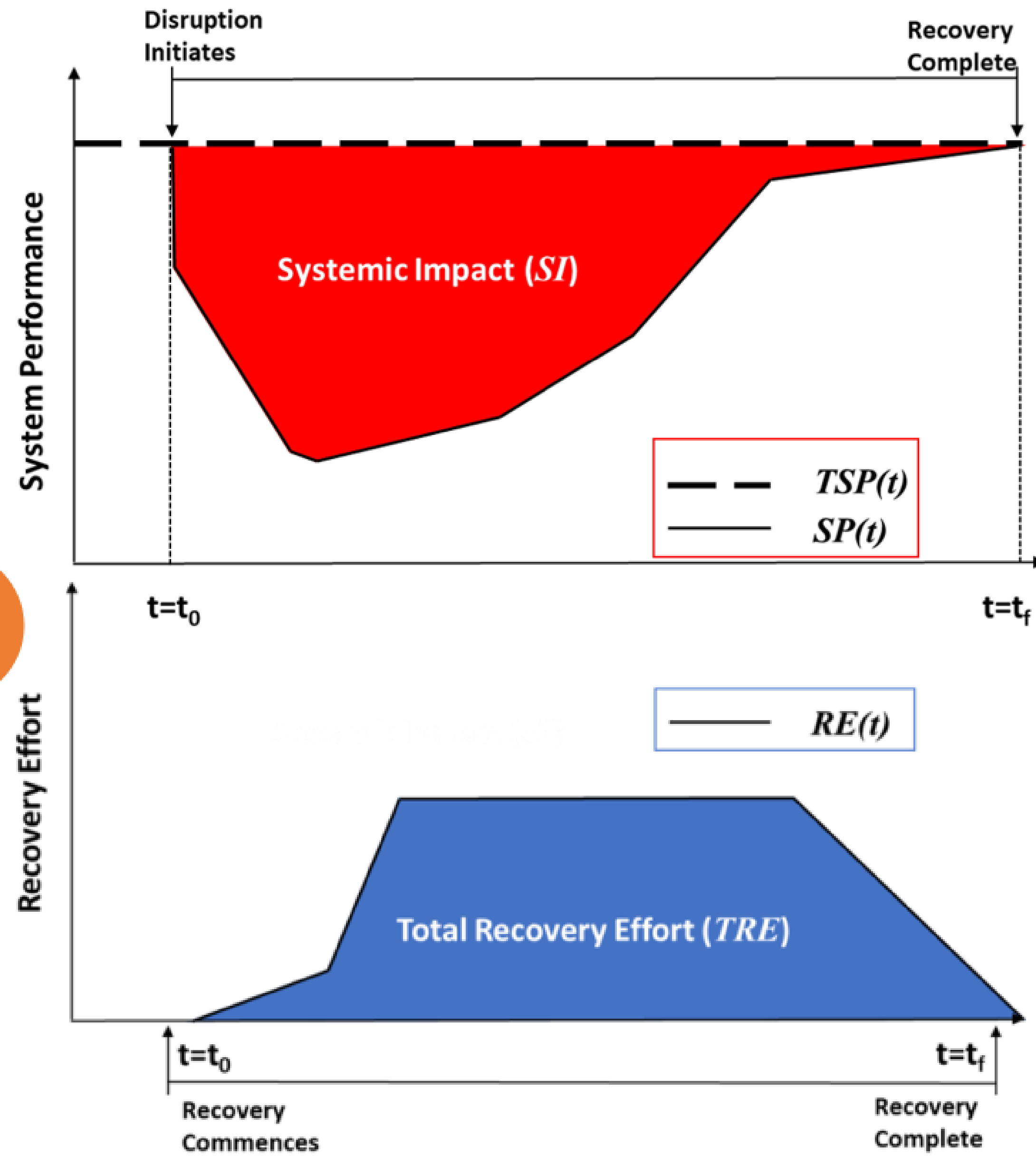
1. Database/ledger -> distributed append-only database/privateDLT
2. Fuse traditionally disparate data, as appropriate, to improve timeliness and Continuity of Knowledge
3. Physical adds to operator protocols, ***boost*** data approaches

Use of Distributed Ledger Technology could improve data efficiency and surety, <u>a rare two-for-one opportunity</u>



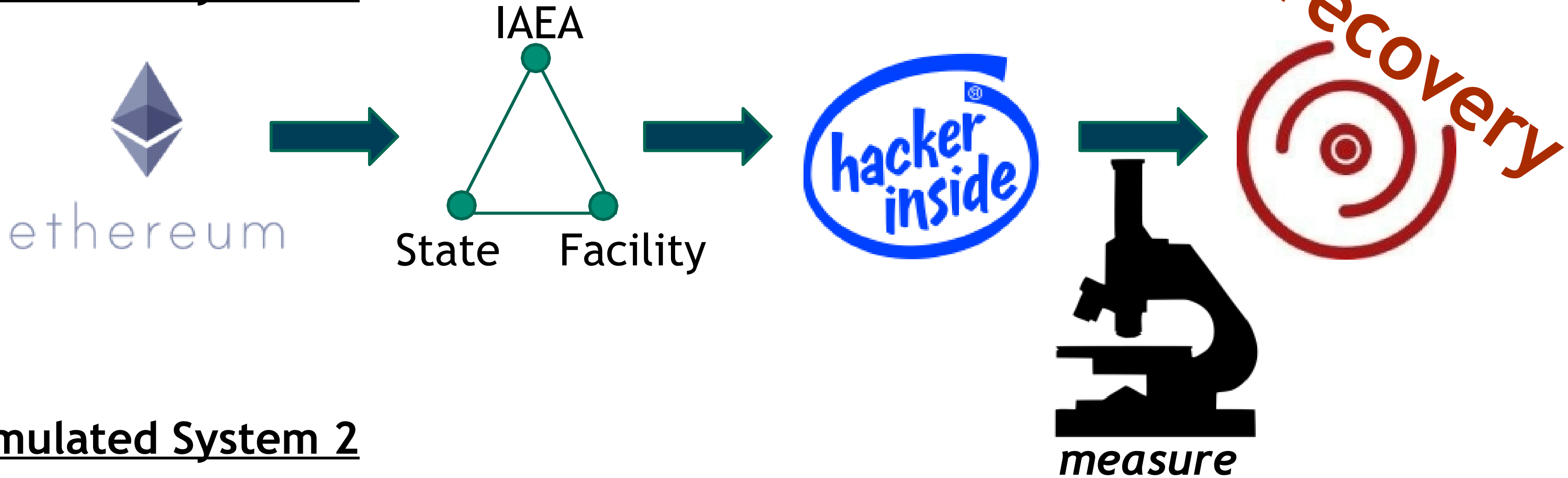
Year 3 Emphasis, Resilience Experimentation



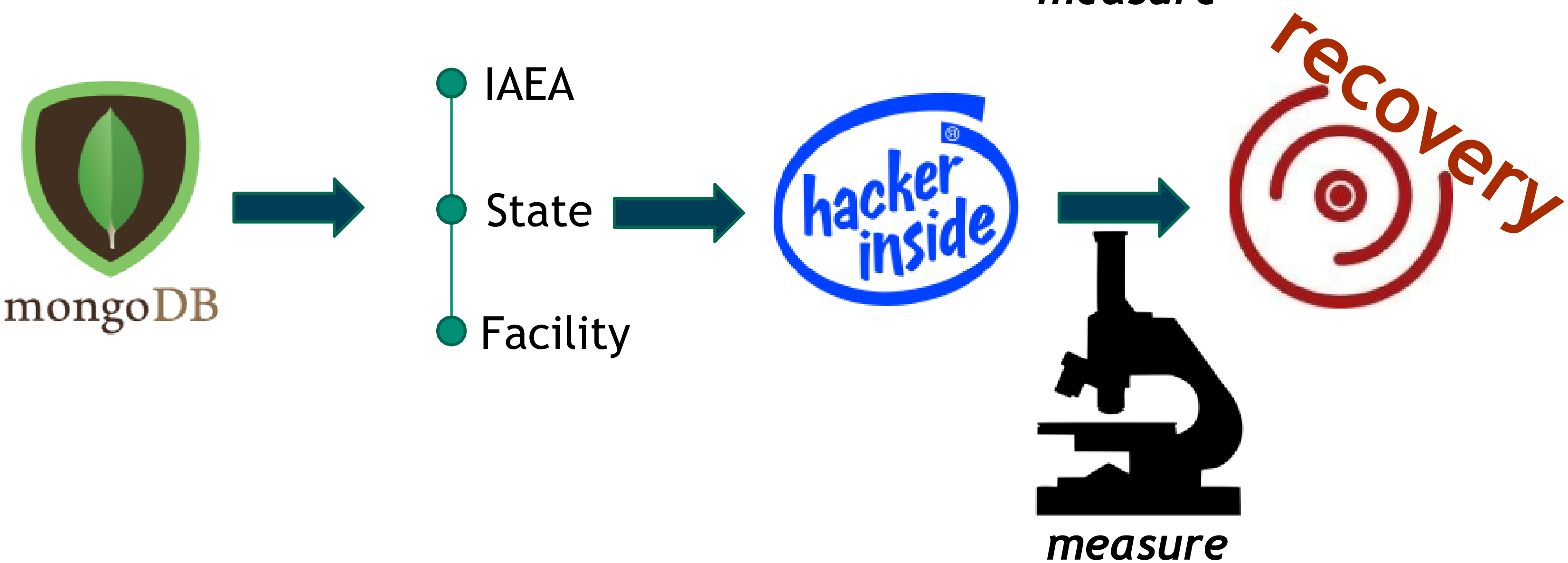


Experimentation Process

Emulated System 1



Emulated System 2



Concrete Resilience Comparisons

$SI_1(t) =$	The confidentiality of data in the system at time t measured by the amount of data that is not accessible by unauthorized parties.
$SI_2(t) =$	The inaccuracy of data in the system at time t measured by the cumulative difference between true known quantities and quantities reported in the system.
$TRE_1(t) =$	The effort to reconcile ledgers at time t measured by the manpower performing a reconciliation task at time t.
$TRE_2(t) =$	The effort to locate a physical asset at time t measured by the manpower performing a location task at time t.
$TRE_3(t) =$	The effort to identify an asset is missing at time t measured by the manpower performing an identification task at time t.

<u>Example Scenarios</u>

Scenario 1: unauthorized data modification to State datastore after already shared with IAEA, resulting in a data discrepancy. Noticed during subsequent report period.

Scenario 2: undetected theft of material from facility, resulting in data discrepancy Noticed at inventory.

Scenario 3: corruption of State datastore via damage to data storage equipment.

Concrete Resilience Comparison (continued)

	DLT-enabled	Traditional Database
R_1^{SI}, i.e. data accuracy	Scenario 1: 3x trad Scenario 2: 3x trad Scenario 3: 1/3, short	Scenario 1: small (< 1%) Scenario 2: small (< 1%) Scenario 3: 1/3, long
R_2^{SI}, i.e. data confidentiality	Not affected	Not affected
R_1^{TRE}, i.e. time reconciling ledgers	Scenario 1: none Scenario 2: same, once noticed Scenario 3: none	Scenario 1: significant Scenario 2: same, once noticed Scenario 3: significant
R_2^{TRE}, i.e. time to locate asset	Scenario 1: long Scenario 2: n/a Scenario 3: n/a	Scenario 1: long Scenario 2: n/a Scenario 3: n/a
R_3^{TRE}, i.e. time to identify as missing	Scenario 1: n/a Scenario 2: long Scenario 3: n/a	Scenario 1: n/a Scenario 2: long Scenario 3: n/a

	DLT-enabled	Traditional Database
SCENARIO: unauthorized data modification	$\frac{.97+1+1+.75+1}{5} = .944$	$\frac{.99+1+.25+.75+1}{5} = .798$
SCENARIO: theft	$\frac{.97+1+.75+1+.25}{5} = .794$	$\frac{.99+1+.75+1+.25}{5} = .798$
SCENARIO: datastore loss	$\frac{.8+1+1+1+1}{5} = .96$	$\frac{.5+1+.5+1+1}{5} = .8$

Multi-Party Computation (MPC) Provides a Means to Share Proprietary or Sensitive Data

Generally missing from the IAEA collection is the plethora of 'big data' being continually generated by the nuclear facility for operator purposes, but this data is considered proprietary by the nuclear facility operators.

Use of Multi-Party Computation (MPC) could obviate the proprietary issue since the operator <u>never reveals the underlying data</u>

The IAEA could have a new stream of otherwise inaccessible nuclear facility operator data to complement typical safeguards data.

This same MPC technology could also allow nuclear facilities with different data sensitivity concerns to share data amongst themselves.

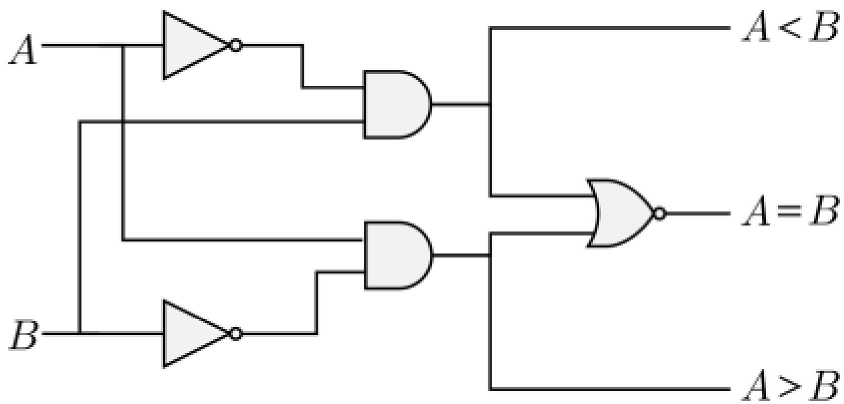
Modality	IAEA Data Sources	Operator Data Sources
Quantitative Sensors	Gamma ray spectrometry (U and Pu isotopics)	Water chemistry (pH, ppm levels, conductivity, hydrogen, oxygen, chloride, fluoride, boric acid concentrations),
	X-ray spectrometry (element identification, container thicknesses)	Primary and secondary loop temperatures, pressures, flow rates, water levels
	Neutron counting (U and Pu amount/enrichment verification)	Accelerometers (vibration FFT)
Operational Signatures	Power monitor (Advanced Thermo-hydraulic Power Monitor)	Ex-core neutron flux (noise shows vibration, phase differences between detectors)
		Reactor power
		Control rod positions
		Steam generator pressures & flow rates
		Valve settings (open/closed)
	Cerenkov radiation viewing	Radiation monitors
		Motor current signature analysis (>350 motors to drive pumps, fans & compressors)
		acoustic emissions monitoring (emitted from equipment and pressure boundaries)
		Odor, burning, fumes
Containment & Surveillance	Camera surveillance	Security cameras
	Load cells (weight measurements)	
	Seal inspection	RFID tracking
	Containment verification (e.g. laser reflectometry)	
Off-site Laboratory	Destructive Assay (alpha, x-ray, gamma, mass spectrometry, etc.)	Personnel radiation monitors
Environmental Sampling	Particles	Gas effluents
Documentation	Inspector reports, Inventory ledger reconciliation	Maintenance reports, INPO/WANO visits, Regulator event notification reports
Design Information	3-D laser range finder	Security personnel

Table 1: Types of data sources typically used by the IAEA for safeguards at nuclear power plants; and typical data sources used by civilian reactor operators.

"Garbled Circuits" (2-party MPC) is Working

The ***CypherCircuit*** Python library has been built and is running on applicable problems.

Simple Comparator circuit



FACILITY

```
circuit = CircuitBoard()
A, B = Wire(circuit), Wire(circuit)
comparator = OneBitComparator(A, B)
circuit.garble()
diagram = circuit.sketch()
encoding = circuit.encode([0, 1])
```

IAEA

```
circuit = CircuitBoard(diagram)
decoding = circuit.decode(encoding)
print(decoding)
```

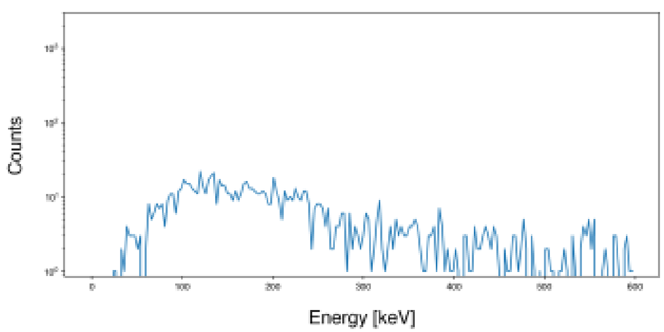
```
Out [1]: [1, 0, 0]
```

- Package emphasizes transparency & clarity of garbled circuit methodology to audiences without cryptographic backgrounds (e.g. a general safeguards "customer")

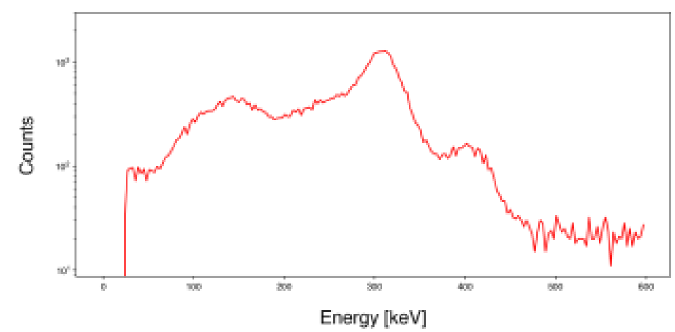
FY20: Using Garbled Circuit on ORNL MUSE Safeguards Data

- We were able to get ORNL concurrence to give Berkeley two months (Feb/Mar 2019) of MINOS/ MUSE radiation detector spectral data (*Jun. 2020*)
  - Feb has one instance of a confirmed Np transport event; March has two Np-Pu events
- "Regions of Interest" were used on the MUSE spectra to hunt for anomalous high counts of notable gamma rays (e.g. 311 keV for Pa-233) compared to normal
  - March MUSE data has so much clutter; will require a more discerning algorithm

MUSE background



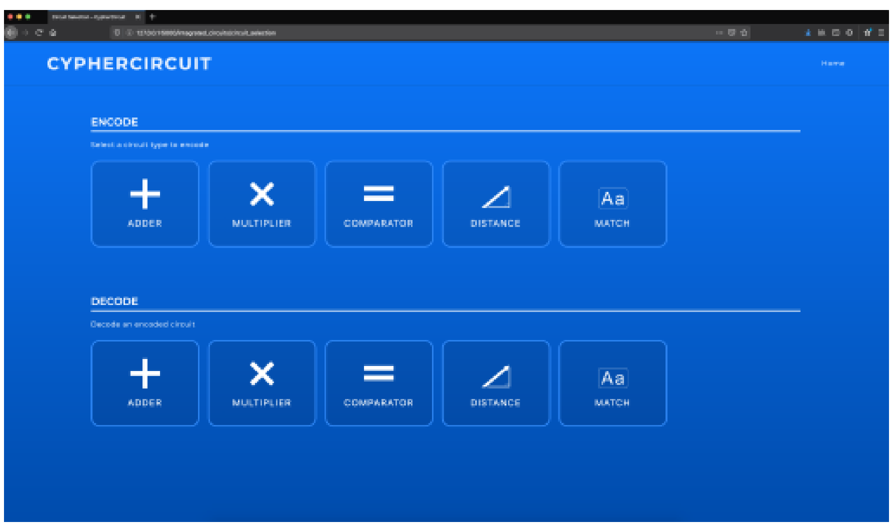
Identified Neptunium Transfer



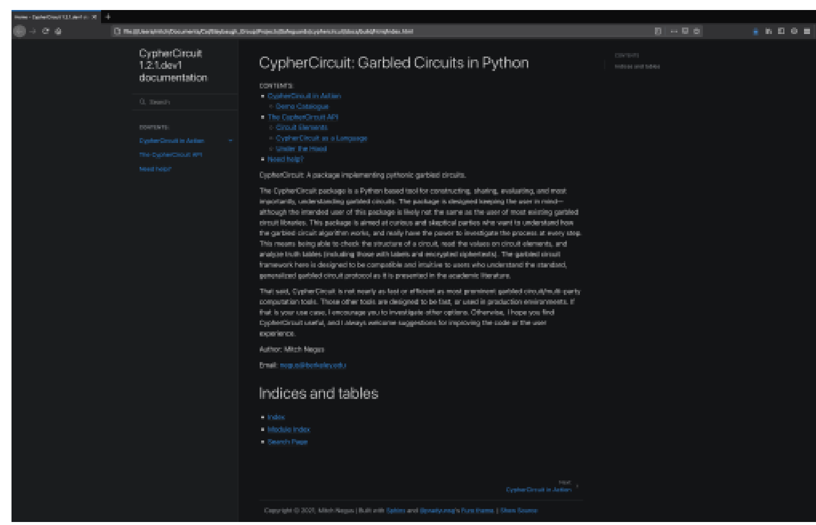
- For Feb., a 10 min sample of radiation spectra collected by the MUSE sensor array was processed by a garbled circuit; complete calculation took ~14.5 hours (*Sep. 2020*)
  - Anomaly detected based on ratio of 311 keV gamma peak to 398 & 415 keV gammas (above)
  - 936,935 gates total; FreeXOR optimization has reduced computation cost by ~13%
  - Substantial quantity of remaining time is Oblivious Transfer (OT); can be significantly improved using OT extension technique (to be implemented by end of FY21)

An Improved, User-Friendly Version of CypherCircuit is Available

- Successfully navigated the Export Control and Intellectual Property offices at Sandia to allow delivery of CypherCircuit to USG and National Labs (*Oct. 2020*)
  - Package can be accessed via GitLab or using distribution ZIP file
    - Includes easy install, full tutorial, comprehensive documentation, and example demos
    - ORNL is expected to access and use the package in the Red/Blue team exercise
  - Network procedure enhanced to facilitate smooth (online) multiparty interaction
  - Code now imitates a complete "language" (*Mar. 2021*)
    - Standard operations (+, -, ×, ÷, >, <) can be specified as code; circuits do not need to be built by hand as individual wires and gates



The *CypherCircuit* user interface



The *CypherCircuit* (Sphinx-based) documentation

Speed Comparisons to State-of-the-Art

- *CypherCircuit* is slow for MUSE data analysis (hours for a solution)
- We are comparing the speed of CypherCircuit with other open-source garbled circuits

(Euclidean Distance)^2 Time, 10 trials [s]1		
Dimensions	Obliv-C*	CypherCircuit
2	3.303	462.4
3	2.922	740.3
4	3.456	919.5
5	3.477	1207.8
10	2.989	2465.3
100	3.466	27406.5
1000	6.88	—

- Timing benchmarks are Dockerized/version-controlled, can be reproduced with minimal effort
- Obliv-C was chosen as one of the “state-of-the-art” codes as it is intended for non-expert users
  - Still required about one week to learn and perform the most basic of implementation tasks
  - *Seems to have difficulty running calculations in quick succession due to network setup

Potential Solutions to Enhance Speed

- OT Extension
- Parallelization (of gate evaluation and/or circuit iteration)
- Backend swap (more Cython, full C/C++ implementation, “state-of-the-art” code as backend?)
- FPGA acceleration?

Future Plans for MPC/Garbled Circuits

Technical challenges

- Implementation is secure for *semi-honest adversary* (follows the protocol, but tries to figure out other party's data)
  - Methods to address *malicious adversary* are known, but even more computationally expensive (e.g. perform zero-knowledge proofs with garbled circuits)

Remaining FY21 Work

- Build garbled circuit to perform grammar compression
  - Operation on compressed data should enable faster anomaly detection with garbled circuits for relevant problems
  - Grammar compression circuit is enabled by new "language" capability of CypherCircuit (can be similarly implemented in Obliv-C or other "language"–like MPC frameworks)
- Shift towards improving the algorithms to operate more efficiently on larger, more complete datasets and using more sophisticated methods
- Conduct "Blue Team/Red Team" exercises
  - Teach red team to build/evaluate garbled circuits using CypherCircuit
  - Provide successively more challenging exercises in garbled circuit anomaly detection (e.g. can they hide an anomaly from the detection algorithm)
- See if we can find the same anomaly in the Np-Pu data (logs) that we found in the MINOS/MUSE time series data of Year 2.

Investigate Integration, the Road Ahead

Investigate potential for integrating software tools

Implementation plan:

- Investigate incorporating grammar compression into a garbled circuit
- Demonstrate the three approaches operating in series on a common dataset

Evaluate the full system

Implementation plan:

- Evaluate the performance of an integrated system, in which the three approaches operate in series on a common dataset
- Integrate a notional disruption scenario into demo, with a stretch goal of applying resilience methodology (stretch justification: involves significant effort to define metrics)

Thank you!