SAND2021-9009C

Proceedings of the INMM & ESARDA Joint Annual Meeting
August 23-26 & August 30-September 1, 2021, Virtual Meeting

DEVELOPMENT OF NOVEL APPROCHES TO ANOMALY DETECTION AND SURETY FOR SAFEGUARDS DATA – YEAR TWO AND THREE RESULTS

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], Daniel Archer[4], Michael Willis[4], James Ghawaly[4], Andrew Nicholson[4]

[1]Sandia National Laboratories, Albuquerque, NM 87185 and Livermore, CA 94550, USA
[2]George Mason University, 4400 University Dr., Fairfax, VA 22030, USA
[3]University of California at Berkeley, 4173 Etcheverry Hall, Berkeley, CA 94720, USA
[4]Oak Ridge National Laboratory, Oak Ridge, TN 37830, USA

ABSTRACT

The first phase of the Novel Approaches to Anomaly Detection and Surety for Safeguards Data project which considers the applicability for international safeguards of three core data analysis and management methods was presented at the Institute for Nuclear Materials Management (INMM) Annual Meeting in 2020. Year One of the project saw three major accomplishments. The first accomplishment was the prioritization and selection of anomaly detection methods to improve and extend the existing Grammar Compression (GC) method. One of the key results has been the development of a new method that combines GC with ensemble learning to perform robust and efficient anomaly detection in time series data. The second accomplishment was the down-selection of technologies and data for the prototype Distributed Ledger Technology (DLT) system. We have introduced and described a framework by which adoption tradeoffs of DLT for improved Continuity of Knowledge are being objectively evaluated. And the third accomplishment was an assessment of the viability of Multi-Party Computation (MPC) via a study of test scenarios to evaluate how easily anomalies in raw data sequences convert through a garbled circuit. These three approaches are natural complements to one another, as DLT and GC-based anomaly detection can be used on traditional safeguards data sources or other available data, and the MPC component allows for exploration of nontraditional data sources in a manner that will protect sensitive operator information. This paper outlines the work performed in Year Two and Three of the project and highlights results on: (1) the development of software tools implementing selected anomaly detection methods to extend and improve the existing GC method, (2) the development and evaluation of a software tool implementing the first version of the prototype DLT system, and (3) the application of the MPC approach to actual safeguards data streams. A significant step towards practical applicability of these technologies to authenticate, protect, and analyze actual safeguards data will be the development of an integrated platform that combines all methods. The paper will conclude with the current status of this development and an outlook on next steps. *SNL is managed and operated by NTESS under DOE NNSA contract DE-NA0003525. SAND2021-1706 A.*

INTRODUCTION

International safeguards authorities increasingly complement physical inspection activities with data that is generated in-between inspection visits, compiled from a variety of open and non-facility related sources, and even operator data, when accessible. The quality of the data as well as the processing tools available have a significant impact on the conclusions that can be drawn from such data. For this reason, the International Atomic Energy Agency (IAEA) is constantly looking to strengthen the surety of safeguards data and to develop new tools to analyze the data at its disposal.

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. SNL Programmatic R&A: 1331307

Sandia National Laboratories (SNL) initiated a project that evaluates the applicability of three data processing tools for use by the IAEA or other safeguards authorities. Our exploration of novel approaches to anomaly detection and surety for safeguards data looked at three core data analysis and management methods: (1) anomaly detection in multivariate safeguards data based on the Grammar Compression (GC) method, (2) development and testing of a novel safeguards data authentication, integration, and analysis workflow on the foundation of Distributed Ledger Technology (DLT), and (3) investigating how operator data could assist in drawing safeguards conclusions in a Multi-Party Computation (MPC) environment.

Results of the Year One of the project were reported on at the International Nuclear Materials Management (INMM) 61st Annual Meeting and the paper showed that all three methods have significant potential for use in international safeguards and if developed with the appropriate framework in mind, might even offer additional synergetic benefits when used in unison [1].

To briefly summarize the impact of each method, we find that by adapting the GC method of anomaly detection to the analysis of safeguards data and testing it on representative data, we can produce a powerful tool for automated discovery of abnormalities in fuel cycle activity, including various types of prohibited events such as material diversion and facility misuse. The new capabilities, along with the included visualization tools, can help the inspectors and analysts to focus their attention on most critical sequences of data, and thus tremendously increase the effectiveness of their work.

In prototyping a DLT for safeguards data provenance, we can have a clear technical path for realizing enhanced confidence and improved efficiency in the safeguards process, as well as a more explicit control over the trust model of safeguards data sharing, leading to increased resiliency, clarity in tolerable risk and finally improved Continuity of Knowledge (CoK).

By implementing MPC, the IAEA or other safeguards authorities can have a new stream of otherwise inaccessible nuclear facility operator data to complement typical safeguards data, due to obviating any proprietary or secrecy concerns. This same MPC technology could also allow nuclear facilities with different data sensitivity concerns to share data amongst themselves, potentially across borders, which would ultimately enable safer operations, more transparency, and therefore better confidence that participating facilities are not being misused.

Taking the step from the evaluation of the three methods towards the development of actual software tools is the logical continuation of the project in Year Two and Three, and the relevant findings are presented in this paper. The successful development of an integrated platform that combines the prototype DLT system and MPC methods with the GC-based anomaly detection software package will represent a significant step towards practical applicability of the proposed technologies to authentication, protection, and analysis of the actual safeguards data.

ANOMALY DETECTION USING GRAMMAR COMPRESSION

One of our goals in this project is to develop a suite of practical methods for effective and efficient detection of anomalies in time-series data obtained from safeguards. The key component of the proposed approach is the cutting-edge method of unsupervised anomaly detection based on GC [2][3]. This method has a number of crucial advantages important for analysis of safeguards data. First, GC scales linearly with data size and therefore is capable of efficiently analyzing a very large amount of data that safeguards generate. Second, GC can be extended to include the capability for detection of correlated anomalies in multivariate data, like those generated by multiple types of

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. SNL Programmatic R&A: 1331307

safeguards sensors. Third, GC can be extended to incorporate ensemble learning for improved robustness against approximation errors.

After investigating key challenges posed by safeguards data, we prioritized anomaly detection methods that would be most valuable in extending and improving the capabilities of the existing GS method. Specifically, we decided to focus our efforts on development, implementation, and testing of three new capabilities: (1) robust, parameter-free anomaly detection by integrating CG with ensemble learning, (2) anomaly detection on extra-long scale, based on efficient, variable-length motif discovery [4], and (3) detection of correlated anomalies in multivariate data.

One of our key achievements in Year One has been the development of a new method that combines GC with *ensemble learning* to perform robust and efficient anomaly detection in time series data. This work resulted in the development of the Ensemble Grammar Induction (EGI) method and the associated software package, whose state-of-the-art performance has been described in previous reports [1][5]. In consequent work, we investigated how the EGI performance depends on the algorithm's hyperparameter (sliding window size). While the incorporation of ensemble learning eliminates dependencies on other parameters (specifically, alphabet size and piecewise-aggregate-approximation size), the accuracy of anomaly detection still depends on how close the hyperparameter value is to the actual anomaly length that is often unknown *a priori*. The performance is evaluated using two metrics: Score (a measure of overlap between the ground truth anomaly and the candidate anomaly found by EGI) and Hit Rate (the fraction of candidate anomalies with non-zero score). Figure 1 shows values of the average Score and Hit Rate, computed over a set of 200 time series for each of 12 open-source datasets from different application areas, as the hyperparameter value varies between one fifth and five times the actual anomaly length. While the performance varies from one dataset to another, the general trend is that the metrics peak at or near the hyperparameter value equal to the anomaly length.

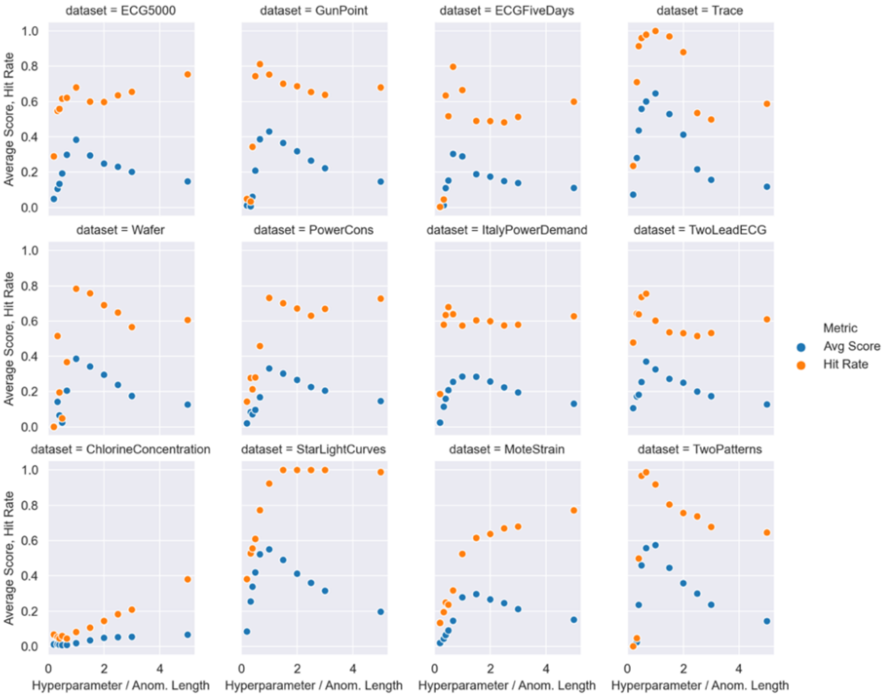


Figure 1. EGI performance metrics as functions of the ratio of the hyperparameter (the sliding window size) to the actual anomaly length.

Furthermore, we developed a new method applicable for *detecting anomalies on extra-long scale* (time series with millions of data points). To achieve this goal, we leveraged a new efficient algorithm for variable-length motif discovery [4]. Motifs are recurrent patterns in a time series. We use motif discovery as a key step in anomaly detection by designating subsequences that contain least number of frequent motifs as anomaly candidates. Specifically, the new method computes a motif correlation density curve (MCDC) whose minima indicate anomaly candidates. Additionally, the length of each anomaly candidate is evaluated by computing the derivative of the MCDC around a minimum point, which does not require any *a priori* information about anomaly length. Figure 2 shows two examples of using the MCDC method for detection of anomalies in time series from different open-source datasets. Example on the left is a time series where the ground truth anomaly is easily identified by inspection. Example on the right is a time series where the ground truth anomaly is not clearly distinguishable from the rest of the data. The MCDC method successfully detects the ground truth anomaly as the top candidate in both cases.

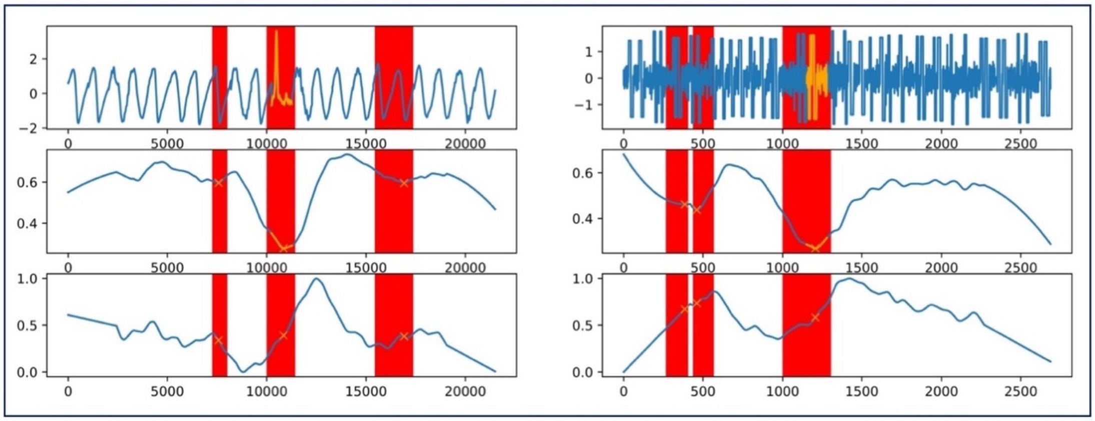


Figure 2. Two examples of anomaly detection using the MCDC method. The top plot is the time series, the middle plot is the MCDC whose minima indicate anomaly candidates, and the bottom plot is the MCDC derivative (normalized) used to determine the lengths of candidate anomalies. Top three anomaly candidates are highlighted by red background color and the ground truth anomaly is outlined in yellow.

A significant portion of our efforts focused on developing a new method for detecting anomalies in multivariate time series data. This 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. Our approach combines GC with deep learning. As shown in Figure 3, at the first step we convert the spectrogram obtained from a radiation detector into a single time series by summing over spectral components with optimal weights selected by a deep neural network (DNN). At the second step, we identify anomaly candidates in the integrated signal by using the EGI method. At the third step, we use a DNN-based time-series classifier to place the temporal segment identified by EGI into one of multiple classes: a normal signal or a signal that corresponds to a particular radioactive material. The classifier is applied to the full spectrum discretized into a number of bands. However, applying EGI to the integrated time series to identify a segment where a potential anomaly is located (at the second step) is crucial for efficiency. We investigated the performance of two time-series classifiers: Long Short-Term Memory Fully Convolutional Network (LSTM-FCN) and Residual Neural Network (ResNet). For testing and evaluation of the new method, we utilized a simulated radiation detection dataset (radDetect) developed by ORNL for open use in the Urban Nuclear Detection Challenge [6] [7] [8]. Numerical experiments with radDetect data show that the best performance is achieved by using ResNet with 34 DNN layers, applied to the full spectrum discretized into 60 bands.

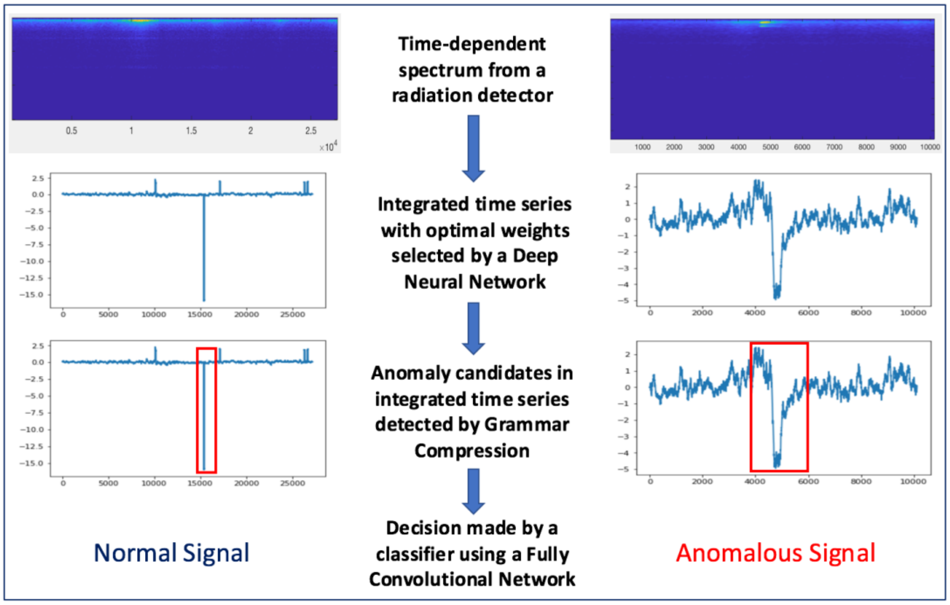

Time-dependent spectrum from a radiation detector
Integrated time series with optimal weights selected by a Deep Neural Network
Anomaly candidates in integrated time series detected by Grammar Compression
Decision made by a classifier using a Fully Convolutional Network
Normal Signal
Anomalous Signal

Figure 3. A schematic of the new method for anomaly detection in multivariate spectral data.

DISTRIBUTED LEDGER TECHNOLOGY FOR DATA PROVENANCE

We have assembled a DLT prototype in order to experimentally explore and characterize tradeoffs between a shared datastore and a more traditional safeguards data flow. Around our prototype, we have built an experimentation infrastructure consisting of:

- **Ground truth engine** - a straightforward state-bearing computer program, which role plays a facility, thereby producing safeguards data events, which are then written to a facility datastore. The facility role playing is from actual safeguards data, i.e., events are produced to mimic a provided real-world dataset, although certain aspects of the real-world facility are estimated/simulated (e.g., the time it would take to recover from a hard drive failure). The ground truth engine is also queryable, which is needed to quantify the accuracy of a datastore at any given point in time.
- **Safeguards reporting process emulator** – a set of scripts that execute notional safeguards reporting actions (e.g., State reporting to IAEA) on a schedule. Additional actions include inventorying/inspection, and datastore reconciliation.
- **Threat emulator** - executes scripted actions to modify the state of one or more components, thereby emulating the after-effects of malicious actions, including:
    - unauthorized data modification (at facility, State, or IAEA)
    - theft/disappearance of material from facility
    - datastore corruption (at facility, state, IAEA), e.g., via damage to equipment
    - compromise of private key material
- **Measurement/analysis engine** – a set of scripts that periodically compare ground truth to the state of the datastore (in the case of DLT) or datastores (in the case of a multiple databases residing at different organizations).

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. SNL Programmatic R&A: 1331307

A more detailed treatment of the resilience analysis methodology we apply in order to compare the performance of our DLT prototype relative to traditional practice is found in our previous paper [1]. In short, resilience is defined as the ability of a system to continue to perform mission essential tasks despite the presence of a disruption or attack. Although related to metrics of effectiveness, these resilience metrics have several advantages: they are quantitative, scenario-based, and temporal-based. Calculating a resilience score amounts to aggregating measured performance over a single scenario. We then calculate resilience scores across a variety of scenarios to evaluate overall resilience.

The performance measures chosen are:

$SI_1(t) =$ 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.

Measures SI directly tie to the mission/system performance of the system, while measures TRE quantify the recovery efforts needed to restore the system to target performance after disruption. The proper approaches for combining measured data into overall resilience scores is still an area of active research. Contemporary practice is to perform a weighted sum, with optional nesting of relationships, in order to derive normalized values that in turn typically yield natural analysis and explanation. For simplicity in this example, we assume that all performance measures are weighted equally, and therefore the overall resilience metric is expressed as:

$$R = \left(1 - \frac{3}{5}\right)\left(\frac{1}{2}R_1^{SI} + \frac{1}{2}R_2^{SI}\right) + \frac{3}{5}\left(\frac{1}{3}R_1^{TRE} + \frac{1}{3}R_2^{TRE} + \frac{1}{3}R_3^{TRE}\right)$$

$$= \frac{1}{5}R_1^{SI} + \frac{1}{5}R_2^{SI} + \frac{1}{5}R_1^{TRE} + \frac{1}{5}R_2^{TRE} + \frac{1}{5}R_3^{TRE}$$

We next provide summary-level data from all executed scenarios, as well as anecdotally explore a scenario in greater detail in order to illustrate the manner in which our resilience results can drive adoption considerations and tradeoffs.

These overall results, while quantitative and founded upon a principled methodology, require analysis and interpretation. It is beneficial that these results are produced in a repeatable and verifiable fashion, as it can move adoption tradeoff discussions from the anecdotal and hypothetical to, instead, whether the emulated threat actions, and measurement methods, are useful for understanding adoption tradeoffs, and can be adjusted/expanded as necessary.

Table 1. Resilience metric calculations

	DLT-enabled	**Traditional Database**
Unauthorized data modification	$\frac{.86+1+1+.93+1}{5} = .958$	$\frac{.92+1+.97+.93+1}{5} = .964$
Theft/disappearance of material	$\frac{.86+1+1+.5+.75}{5} = .822$	$\frac{.92+1+.8+.5+.85}{5} = .814$
Data corruption, e.g., damaged storage	$\frac{1+1+1+1+1}{5} = 1$	$\frac{.8+1+.5+1+1}{5} = .86$
Compromise of private key material	$\frac{1+.66+1+1+1}{5} = .932$	$\frac{1+.66+1+1+1}{5} = .932$

The data shown in Table 1 suggests that a DLT approach is no less resilient to a wide variety of attacks than a traditional database, and (likely unsurprisingly) has increased resilience to attacks related to data tampering, due to audit-friendly data trails, and the lack of a need to reconcile databases between organizations. In turn, this result suggests that the potential improvement in resilience is worth weighing against the cost of adoption.

We conclude this section with a more in-depth analysis of the data around a single scenario (Table 2), to more concretely illustrate the various experimentation components (described earlier) in action.

Table 2. Single scenario analysis

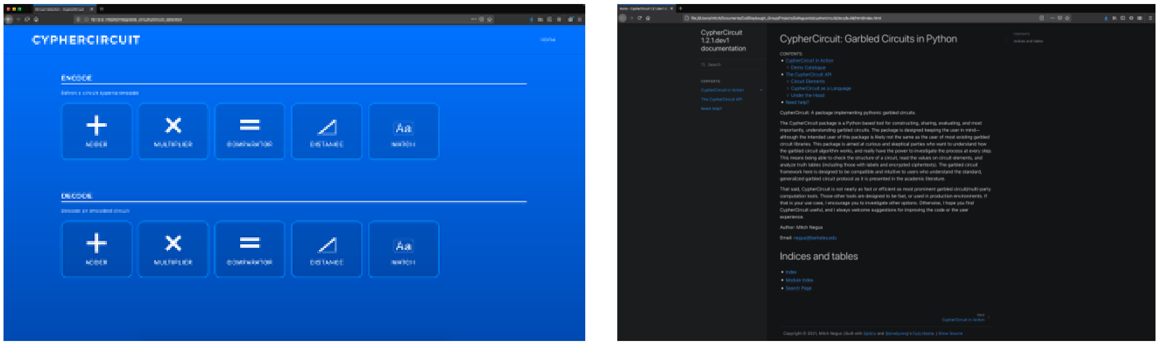
SCENARIO: Unauthorized data modification to state ledger after data already shared w/ IAEA, noticed upon next facility report to state	**DLT-enabled**	**Traditional database**
R_1^{SI}, **i.e., data accuracy**	Until mitigated, all three sites have inaccurate data (0.86)	Limited to SSAC (0.92)
R_2^{SI}, **i.e., data confidentiality**	Not violated in scenario (1.0)	Not violated in scenario (1.0)
R_1^{TRE}, **i.e., time reconciling ledgers**	None, the ledger is identical across organizations (1.0)	Discrepancy identified in one day, via ledger compare (0.97)
R_2^{TRE}, **i.e., time to locate asset**	No material is missing, facility corroborates in two days (0.93)	No material is missing, facility corroborates in two days (0.93)
R_3^{TRE}, **i.e., time to identify as missing**	No material missing in scenario (1.0)	No material missing in scenario (1.0)

While the evaluation of our DLT prototype has been very useful for our project and our sponsor, we hope that the methodology and experimental infrastructure outlined here helps set a precedent for a quantitative and principled approach to adoption of future enabling technologies in high consequence mission domains.

MULTI-PARTY COMPUTATION FOR DATA PRIVACY

MPC allows for parties with sensitive data to contribute to a desired calculation (i.e., a function output based on one or more inputs of data streams), but without revealing the underlying raw, sensitive data to any other party. In essence, MPC uses cryptography to tokenize parties' data such that the tokenized data is unrecognizable by any party yet is useable by the cryptographic function. For our work, we utilize Garbled Circuits [1], a form of MPC that is useful for two-party scenarios. A garbled circuit implementation package called *CypherCircuit* was developed that includes:

- An easy install, full tutorial, comprehensive documentation, and example demos
- Network procedure enhanced to facilitate smooth (online) multiparty interaction
- Standard operations (+, -, ×, ÷, >, <) can be specified as code; circuits do not need to be built by hand as individual wires and gates



CypherCircuit: Garbled Circuits in Python

CONTENTS:

- CypherCircuit in Action
  - Demo Catalogue
- The CypherCircuit API
  - Circuit Elements
  - CypherCircuit as a Language
  - Under the Hood
- Need help?

CypherCircuit: A package implementing pythonic garbled circuits.

The CypherCircuit package is a Python based tool for constructing, sharing, evaluating, and most importantly, understanding garbled circuits. The package is designed keeping the user in mind—although the intended user of this package is likely not the same as the user of most existing garbled circuit libraries. This package is aimed at curious and skeptical parties who want to understand how the garbled circuit algorithm works, and really have the power to investigate the process at every step. This means being able to check the structure of a circuit, read the values on circuit elements, and analyze truth tables (including those with labels and encrypted ciphertexts). The garbled circuit framework here is designed to be compatible and intuitive to users who understand the standard, generalized garbled circuit protocol as it is presented in the academic literature.

That said, CypherCircuit is not nearly as fast or efficient as most prominent garbled circuit/multi-party computation tools. Those other tools are designed to be fast, or used in production environments. If that is your use case, I encourage you to investigate other options. Otherwise, I hope you find CypherCircuit useful, and I always welcome suggestions for improving the code or the user experience.

Author: Mitch Negus

Email: negus@berkeley.edu

Indices and tables

- Index
- Module Index
- Search Page

Next
CypherCircuit in Action

Copyright © 2021, Mitch Negus | Built with Sphinx and @pradyunsg's Furo theme. | Show Source

Figure 4. The *CypherCircuit* user interface (left) and the *CypherCircuit* (Sphinx-based) documentation (right)

A 10-minute sample of radiation spectra collected by the MUSE [9] [10] sensor array at Oak Ridge National Laboratory (ORNL) was processed by a garbled circuit built with our *CypherCircuit* Python library. The complete calculation took ~14.5 hours (Sep. 2020). Anomalies were detected based on the ratio of 311 keV gamma peak to the 398 keV peak, and the 415 keV gammas were the identified anomaly in the data, as shown in Figure 5.

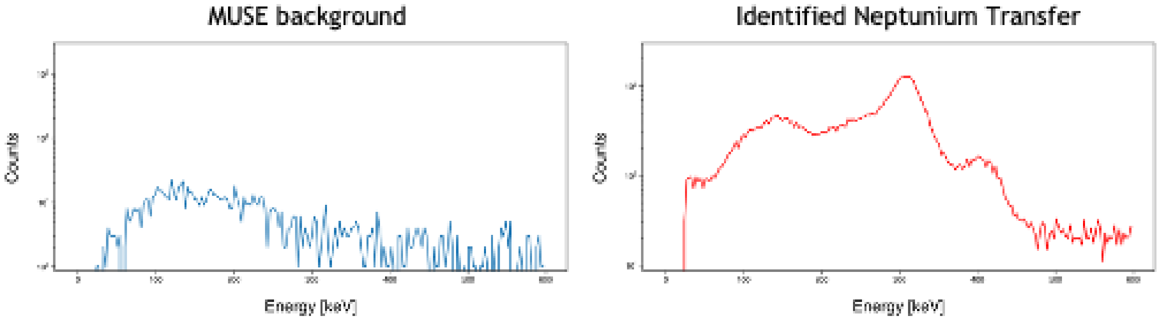


Figure 5. MUSE background and identified Neptunium transfer

Our Garbled Circuit required 936,935 gates total, and subsequent FreeXOR optimization has reduced computation cost by ~13%. A substantial quantity of time cost is Oblivious Transfer (OT), which is used to provide other party random tokens, but can be significantly improved using an OT extension technique.
Since *CypherCircuit* is slow for MUSE data analysis (hours for a solution), we are comparing the speed of *CypherCircuit* with other open-source garbled circuits. The openly available code *Obliv-C* was chosen as one of the “state-of-the-art” codes as it is intended for non-expert users, and some speed comparison tests are shown in Table 3:

Table 3. Garbled circuit speed comparison tests

Euclidean Distance Time (sec) over 10 trials		
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	—

As observed, *CypherCircuit* could certainly improve its speed by orders of magnitude, if the underlying approach of *Obliv-C* can be utilized for our purposes. Other possibilities to increase the speed of *CypherCircuit* include:

- An Oblivious Transfer (OT) Extension to be added to *CypherCircuit*
- 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

Our current work is towards building a garbled circuit implementation of GC as the anomaly detection protocol. When combined with this project's related work on GC for safeguards datasets, enabling garbled circuit based-GC would provide a privacy-preserving means for evaluating large, cyclical time series.

Our GC implementation is being built using the *Obliv-C* framework discussed above. The general structure of this implementation is complete and is based on an equivalent computation structure (implementing standard, non-oblivious GC) that has been developed in *C* to serve as a general template. The algorithm suffers from the fact, however, that the efficiency gains realized by GC in storage of compressed grammar structures are challenging to implement in garbled circuits. These data storage structures rely on efficient mappings that do not need to be traversed in their entirety upon data insertion and data access; however, to preserve privacy, a garbled circuit would require that these operations inspect every data location to avoid yielding information regarding the subject of the operation. To alleviate this dramatic efficiency penalty, our current work focuses on understanding whether other privacy-preserving operations (not garbled-circuit-based methods) can be included in the design. These might include developments like Oblivious Random-Access Memory.

CONCLUSIONS

The work described in this paper represents the second and third phase of the Novel Approaches to Anomaly Detection and Surety for Safeguards Data project which investigates the applicability of three core data analysis and management methods for international safeguards. We continued the work on developing, implementing, and testing new methods that extend and improve the capabilities of the existing GC method. Our key results in this area have been the development of the EGI method that combines GC with ensemble learning for robust and efficient anomaly detection in time series data [5], the analysis of how the EGI performance depends on the algorithm's hyperparameter, the development of a new motif-based method applicable for detecting anomalies on extra-long scale, and the development of a new method that combines GC with deep learning for anomaly detection in multivariate time series data such as time-dependent spectra.

We have evaluated our DLT prototype via a resilience methodology, whereby the data suggests that a DLT indeed yields data integrity improvements over current practice, in the sense that data tampering is more detectable and less localized, likely saving time and recovery effort in reconciling disparate ledgers across facility/State/IAEA boundaries. We hope that our case study also serves as an example for quantitative cyber security experimentation that will be useful in other application domains.

We have successfully identified radiation anomalies in MUSE spectra using garbled circuits, an MPC cryptographic approach. The open-source code *Obliv-C* being much faster, we have used this code both as a speed benchmark for our Python-based code (*CypherCircuit*), as well as the foundation for creating a garbled circuit for grammar compression-based anomaly detection. There are other efficiency gains that could be implemented, including parallelization, writing portions of *CypherCircuit* in C/C++, or hardware acceleration such as using a FPGA. However, for this project we will not pursue these enhancements, but rather will focus on the creation of a working garbled circuit of GC for safeguards-relevant anomaly detection.

REFERENCES

[1] N. Peter-Stein, D. Farley, C. Brif, N. Pattengale, C. Zimmerman, M. Galiardi, Y. Gao, J. Lin, M. Negus and R. Slaybaugh, "Development of Novel Approaches to Anomaly Detection and Surety for Safeguards Data – Year One Results", 61st Annual Meeting of the Institute for Nuclear Materials Management (2020).

[2] P. Senin, J. Lin, X. Wang, T. Oates, S. Gandhi, A. P. Boedihardjo, C. Chen, and S. Frankenstein, "Time series anomaly discovery with grammar-based compression", Proc. 18th International Conference on Extending Database Technology (EDBT), pp. 481–492 (2015).

[3] P. Senin, J. Lin, X. Wang, T. Oates, S. Gandhi, A. P. Boedihardjo, C. Chen, and S. Frankenstein, "GrammarViz 3.0: Interactive Discovery of Variable-Length Time Series Patterns", ACM Transactions on Knowledge Discovery from Data, 12 (1), Article 10 (2018).

[4] Y. Gao and J. Lin, "HIME: discovering variable-length motifs in large-scale time series", Knowledge and Information Systems, 61, pp. 513–542 (2019).

[5] Y. Gao, J. Lin, and C. Brif, "Ensemble Grammar Induction For Detecting Anomalies in Time Series", Proc. 23rd International Conference on Extending Database Technology (EDBT), pp. 85–96 (2020).

[6] Topcoder Challenge: Detecting Radiological Threats in Urban Areas.

[7] J. Ghawaly, A. Nicholson, D. Peplow, Douglas, C. Anderson-Cook, K. Myers, D. Archer, M. Willis, B. Quiter, "Data for training and testing radiation detection algorithms in an urban environment", Scientific Data, Nature (2020).

[8] A. Nicholson, D. Peplow, J. Ghawaly, M. Willis, D. Archer, Daniel, "Generation of Synthetic Data for a Radiation Detection Algorithm Competition", IEEE Transactions on Nuclear Science (2020).

[9] A. Nicholson, D. Archer, I. Garishvili, I. Stewart, M. Willis, Michael, "Characterization of gamma-ray background outside of the High Flux Isotope Reactor", Journal of Radioanalytical and Nuclear Chemistry, volume 318 (2018).

[10] A. Nicholson, J. Ghawaly, I. Stewart, M. Willis, R. Hunley, A. Rowe, Andrew J, D. Archer, Daniel, "Data Collection and Fusion of Persistent Heterogeneous Radiation Detector Network", 60th Annual Meeting of the Institute for Nuclear Materials Management (2019).

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. SNL Programmatic R&A: 1331307