Proceedings of the INMM 61st Annual Meeting
July 12-16, 2020, Baltimore, MD, USA

SAND2020-5665C

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

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

[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

ABSTRACT

The Novel Approaches to Anomaly Detection and Surety for Safeguards Data project, which was introduced at the Institute for Nuclear Materials Management annual meeting in 2019, investigates three core data analysis and management methods and their applicability for international safeguards: Distributed Ledger Technology (DLT) for data authentication, anomaly detection based on Grammar Compression (GC), and how operator data could assist in drawing safeguards conclusions in a Multi-Party Computation (MPC) environment. This paper outlines the work performed in Year One of the project, highlighting results and their impact on the continuation of the tasks. For DLT technologies, an experimental comparison of current safeguards practice versus a DLT-backed prototype is described. For GC-based anomaly detection, we present the investigation of new methods with improved performance based on ensemble learning and variable-length motif discovery. With regards to MPC-based data protection, the viability of applying the method for anomaly detection is analyzed. These three main initiatives are being developed in parallel, with constant cross-effort interactions, and the joint goal of creating an integrated demonstration platform that combines a DLT prototype for safeguards data authentication, GC-based anomaly detection, and MPC-based data integration for sensitive facility information. An outlook on the remaining work for Year Two of the project concludes the paper. *SNL is managed and operated by NTESS under DOE NNSA contract DE-NA0003525. SAND No. 2020-1444.*

INTRODUCTION

IAEA inspectors rely heavily on data coming from the safeguards equipment permanently installed at facilities in order to draw safeguards conclusions. So by nature, nuclear safeguards is a data-rich field and as such holds high potential for the application of modern data analytics techniques.

However, while such analysis methods are used in other areas, they are not yet sufficiently advanced for safeguards use. The Sandia National Laboratories (SNL) team is part of a multidisciplinary effort at three national laboratories to advance a suite of data analytic capabilities to support safeguards activities at declared facilities [1]. In particular, the SNL team focuses on investigation of three core data analysis and management methods and their applicability for international safeguards: (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. In the Year One of the project, our team has completed several deliverables, including: (1) a report on

proposed safeguards use cases, (2) prioritization and selection of anomaly detection methods to improve and extend the existing GC method, (3) down-selection of technologies and data for prototype DLT system, and (4) an assessment of MPC viability via a study of test scenarios with known anomalies to evaluate how easily anomalies in raw data sequences convert through a garbled circuit.

Our current work in Year Two of the project focuses on several objectives: (1) development of software tools implementing selected anomaly detection methods to extend and improve the existing GC method, (2) development of a software tool implementing the first version of the prototype DLT system, and (3) application of the MPC approach to actual safeguards data streams.

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 safeguards sensors. Third, GC can be extended to incorporate ensemble learning for improved robustness against approximation errors. Finally, GC employs unsupervised learning and hence does not require labeled training datasets that are lacking in the safeguards area.

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 four new capabilities: (1) robust, parameter-free anomaly detection by integrating CG with ensemble learning, (2) anomaly detection on extra-long scale (time series with millions of data points), based on efficient, variable-length motif discovery [4], (3) anomaly detection in video data, and (4) 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 [5]. The standard GC requires preselecting values for at least two parameters at the discretization step. How to choose these parameter values properly is still an open problem. Instead of using a particular combination of parameter values for GC-based anomaly detection, our new method generates the final result based on a set of results obtained using an ensemble of GC algorithm executions with different parameter values. Numerical experiments performed on datasets with known ground truth showed that ensemble GC can outperform existing GC-based approaches with different criteria for selection of parameter values. We also showed that ensemble GC, which has a linear time complexity with respect to the data size, can achieve performance similar to that of *discord discovery* [6][7], the state-of-the-art distance-based anomaly detection method that has a quadratic time complexity.

To evaluate the performance of ensemble GC against baseline methods, we used six open-source datasets from different application areas (medicine, manufacturing, 3D motion tracking, synthetic sensor data, and astronomy). For each dataset, we generated 25 time series with

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: 1139045

randomly planted anomalous segments. We compared ensemble GC against four baseline methods: GC with randomly selected parameter values (GC-Random), GC with fixed parameter values (GC-Fix), GC with optimally selected parameter values (GC-Select), and a state-of-the-art implementation of discord discovery [7]. Each of the tested methods returned top-3 ranked anomaly candidates for each time series. We evaluated the performance of each method by using a quantity called Score that quantified the overlap between the discovered anomalies and the ground truth (the planted anomaly). Table 1 shows Score values averaged over the 25 time series for each of the six datasets. Ensemble GC has the highest average Score for four out of six datasets and discord discovery has the highest average Score for two out of six datasets. Figure 1 shows Score values for all 25 time series for three of the datasets. A point in the lower triangle corresponds to a superior performance by ensemble GC compared to the baseline method. While ensemble GC and discord discovery exhibit comparable performance, the former is much more efficient since its time complexity scales linearly with respect to the data size.

Table 1. Performance evaluation results: Score averaged over 25 time series with randomly planted anomalies, for Ensemble GC and four baseline methods.

Dataset	Ensemble GC	GC-Random	GC-Fix	GC-Select	Discord
TwoLeadECG	0.3951	0.2873	0.0629	0.1663	**0.4931**
ECGFiveDay	0.3903	0.2988	0.2671	0.1050	**0.4794**
GunPoint	**0.4728**	0.3715	0.2411	0.0560	0.4000
Wafer	**0.3179**	0.2126	0.1382	0.2480	0.3090
Trace	**0.5718**	0.2022	0.3601	0.3408	0.2816
StarLightCurve	**0.9369**	0.6930	0.5301	0.8759	0.9161

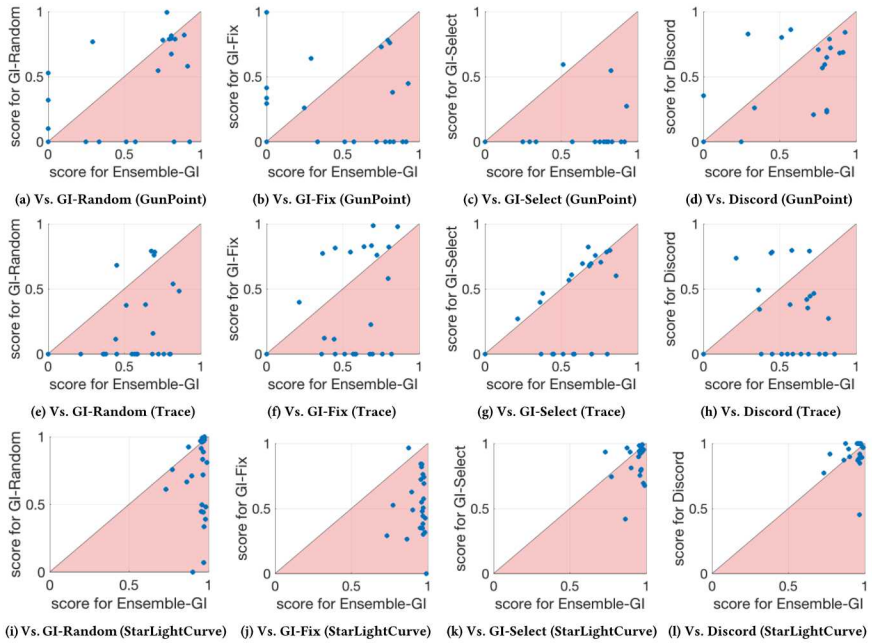


Figure 1. Performance evaluation results: Score values (blue points) for 25 time series with randomly planted anomalies for three datasets (GunPoint, Trace, and StarLightCurve). A point in the lower triangle corresponds to a superior performance by ensemble GC compared to the baseline method.

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: 1139045

Another direction of research that we are currently pursuing is detection of anomalies on extra-long scale (time series with millions of data points). GC is a "greedy" algorithm that tends to focus on variations that occur on a short time scale. Therefore, to detect anomalies on extra-long scale, we are exploring an idea that leverages a new variable-length motif discovery algorithm, Hierarchy-based Motif Enumeration (HIME) [4]. Motifs are recurrent patterns in a time series. Our idea is that motif discovery can be used as a key step in anomaly detection — subsequences that contain least number of frequent motifs are anomaly candidates.

We are also investigating the use of GC-based approaches to detect anomalies in video data. A straightforward approach to analyzing video data is to consider each pixel as a separate time series. However, this would be very inefficient. Instead, we are exploring an idea based on tracking of moving objects. First, we can use an object tracking algorithm to extract trajectories of all moving objects. Second, we can use one of the newly developed GC-based methods (e.g., ensemble GC or the motif-based method outlined above) to detect anomalous trajectories.

DISTRIBUTED LEDGER TECHNOLOGY FOR DATA PROVENANCE

An aspect of our project is prototyping and evaluating distributed ledger technology (DLT) as an enabler for improved data security in safeguards, both international and domestic. DLT is an attractive paradigm to strengthen Continuity of Knowledge (CoK) [8] by eliminating/consolidating redundant material provenance data at the multiple levels (facility, state, international) of Nuclear Material Control and Accountability (termed NMC&A in domestic safeguards, and referred to as Nuclear Material Accounting and Control, NMAC, in international safeguards).There are, however, many practical barriers impeding the adoption of DLT: some technical, some policy, and some based on perception. We introduce and describe a framework by which adoption tradeoffs are being objectively evaluated.

Multiple studies have proposed and investigated goodness of fit of DLT as a technical enabler for achieving shared safeguards ledgers [9] [10]. DLT has been demonstrated as a compelling option in other domains (e.g. the global diamond supply chain) for accumulating product provenance records that are particularly tamper resistant [11]. The sensitivity and confidentiality of safeguards data impedes wholesale application of the techniques applied in more transparent data ecosystems (e.g. supply chains), and the type of DLT required in safeguards are so-called *permissioned DLTs*. Permissioned DLTs have weaker forgery and tamper resistance than public DLTs, and have security properties comparable to distributed, append-only databases. That is, permissioned DLTs afford data access control, and differential privacy [12], and it is straightforward to support a facility submitting data to a permissioned DLT that is 1) visible only to the State and confidential from all other facilities, and 2) not be decipherable by the International Atomic Energy Agency (IAEA) until the State deems it allowable.

To illustrate the DLT approach at a high level, consider the two alternate data topologies in Figure 2. In traditional practice (top of Figure 2), facilities submit NMC&A data to their State authority, typically within 15 days of the end of a month. The State rolls up and investigates facility reports, and submits to the IAEA, typically within 30 days of the end of a month [13]. In the DLT, or shared data, approach (right of Figure 2), the same data (and business processes, and timelines) may be used, however there is only ever a single copy of the data – data exporting and importing is obviated. Straightforward application of public key cryptography can protect

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: 1139045

the confidentiality of facility-submitted data to only their State, until the State grants visibility to the IAEA.

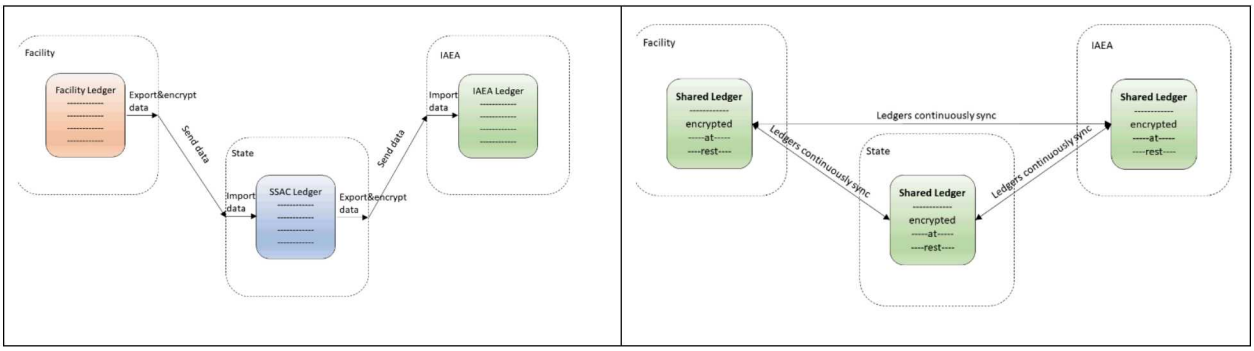


Figure 2. Traditional safeguards practice where organizations export and send data (left) versus a shared ledger concept where all organizations possess a replicated data store (right). Note that in the shared ledger concept, data is encrypted at rest, and decryption keys are only distributed to parties as needed. This enables, e.g. the IAEA, to possess shared data (in an inscrutable form) until deemed releasable by the State.

We wish to evaluate the performance of a DLT versus traditional practice. Previous studies have examined metrics of effectiveness of traditional safeguards practice [14][15]. However, in this work will evaluate using the metrics of resilience [16][17]. There are many definitions of resilience, but we loosely define resilience 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, the resilience metrics presented here have several advantages: they are quantitative, scenario-based, and temporal-based.

Resilience calculations [16] take into account two factors, Systemic Impact (SI), and Total Recovery Effort (TRE), which are integral quantities (i.e. over time) for a given disruption scenario, Figure 3.

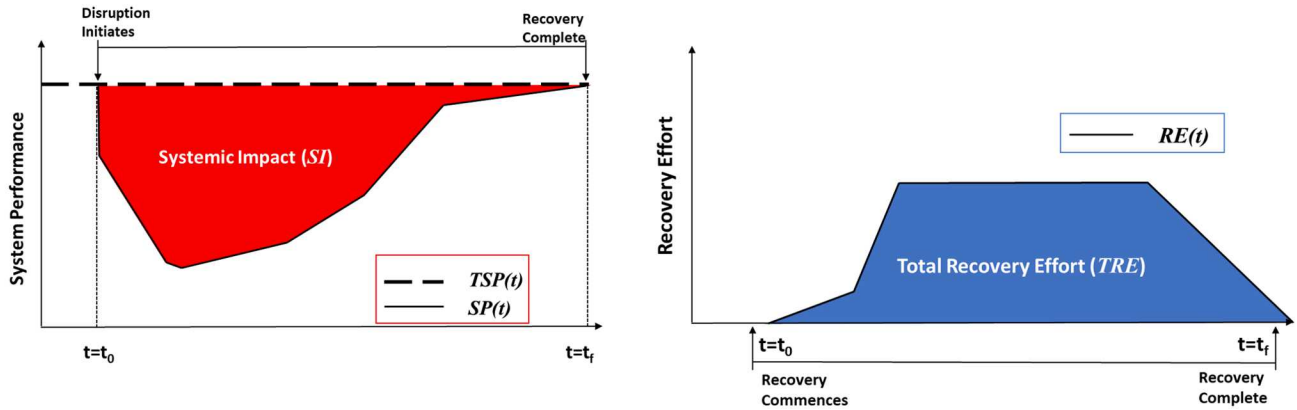


Figure 3. Systemic Impact (SI) and Total Recovery Effort (TRE).

SI and TRE provide different measures of the effect of an impediment a system, and to consider them together provides a description of the total impact of the impediment on the system. Equation (1) combines quantities into a single resilience measure, R, by weighting SI and TRE

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: 1139045

with the factor α. A smaller α value indicates that system performance is more important than usage of response resources; a larger α value indicates that response resources are limited and of greater importance than system performance. The structure of R is well-suited for informing trade-offs between performance and response efforts/costs.

$$R = (1-\alpha)R^{SI} + \alpha R^{TRE} \quad (1)$$

where R^{SI} and R^{TRE} are derived by combining observations (derivation details omitted for space):

$SP_1(t) =$	Confidentiality of data in the system at time t measured by the amount of data that is not accessible by unauthorized parties.
$SP_2(t) =$	Inaccuracy of data in the system at time t measured by the cumulative difference between true known quantities and quantities reported in the system.
$RE_1(t) =$	Effort to reconcile ledgers at time t measured by the manpower performing a reconciliation task at time t.
$RE_2(t) =$	Effort to locate a physical asset at time t measured by the manpower performing a location task at time t.
$RE_3(t) =$	Effort to identify an asset is missing at time t measured by the manpower performing an identification task at time t.

This framework is currently being applied in an isolated experimentation network to evaluate resilience of DLT versus current safeguards practice across a variety of disruption scenarios, both cyber and physical. The disruption scenarios are executed both in our DLT prototype and a traditional architecture, with automated performance measurement, thereby affording repeatability and further experimental exploration of parameter spaces as appropriate.

MULTI-PARTY COMPUTATION FOR DATA PRIVACY

The variety and volume of potentially safeguards-relevant data generated by a nuclear facility is generally not available to the IAEA. MPC is a method that would allow nuclear facility operators to contribute their data into a combined result with other parties' data, yet never exposing the underlying raw data.

While MPC in general can include the participation of an unlimited number of parties, for this effort we are only considering two parties: the IAEA and the State facility. For such a two-party application, the MPC approach of "garbled circuits" makes sense. With garbled circuits, the two participants each take on a role in generating or evaluating the garbled circuit, and then sharing the combined result with each other (although sharing the result is not required). Some of the advantages of the two-party garbled circuit approach are:

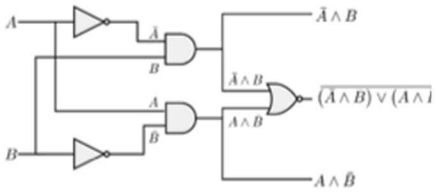
- Straight-forward implementation in Python or other programming language
- Two-party input and roles
- On-line, (near) real-time results
- Fast calculations possible, compared with other MPC implementations

In the garbled circuit protocol that we use [18], a digital circuit is made that represents the desired function to be solved based on input data. As an example, consider the simple comparator digital circuit shown in Figure 4. The circuit simply determines whether party A's or party B's values are greater, less than, or equal to each other, and consists of AND, NOR and NOT gates. The associated truth table is also shown, and, as can be seen, does indeed provide the desired functional result. We use Python to code our algorithms, and the simple script for the comparator is also shown in Figure 4. We emphasize that all of the functional details shown in Figure 4 are not secret and are openly available to all involved parties. We would also like to

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: 1139045

note that NOT gates are essentially "free" as they require no computation, and efficiency gains are possible by using XOR gates as much as possible [18].

Comparator Digital Circuit



(a)

Comparator Truth Table

A	B	A < B	A > B	A = B
0	0	0	0	1
0	1	1	0	0
1	0	0	1	0
1	1	0	0	1

(b)

Comparator Python Code

```
def comparator(A, B):
    if A < B:
        return 1, 0, 0
    elif A > B:
        return 0, 1, 0
    else:
        return 0, 0, 1
```

(c)

Figure 4. Illustration of (a) comparator digital circuit; (b) comparator truth table; and (c) comparator Python code.

Rather than describe the garbled circuit protocol in mathematical form, we will illustrate the garbling process with a pictorial approach, as shown in Figure 5. We show the garbling of just an AND gate () in Figure 5. The top row shows how the Generator ("Alice") generates and assigns random tokens for every '0' and '1' possibility of the AND gate truth table. Here, we use random colors for illustration, but in practice the tokens would be N-bit-length random numbers. The random tokens for inputs A and B (e.g. for Alice and Bob, respectively), are used to encrypt the random token for the gate outputs (column C). Here, the colors would mix into a new color, completely obfuscating the original colors of columns A, B and C. In practice the two N-bit-length random number tokens of columns A and B would be used as numeric keys to encrypt the column C token, again, obfuscating the values of the A, B, and C tokens. Finally, the rows of the truth table are randomly shuffled to erase any information to be gained from knowing the original order of the rows. The Generator then sends to the Evaluator the garbled truth table, as well as the token representing her actual value. Additionally, the Generator can send the Evaluator the two possible tokens for the output (column C), although this is not required (alternatively, Bob could ask Alice which of his decrypted tokens is the correct one).

Following receipt of the garbled truth table and Alice's input token (and likely the tokens for the two possible output tokens), the Evaluator ("Bob") still needs the token corresponding to his input value. He cannot ask the Generator ("Alice") directly for his token corresponding to his '0' or '1', since that would inform Alice of his actual value. One method to provide the Evaluator with his token without compromising his raw data is to use the Oblivious Transfer (OT) protocol [19]. As illustrated in the second row of Figure 5, in 1-out-of-2 OT [20], the Generator places both of the Evaluator's input token possibilities into the possession of a trusted third party (which is likely a secure machine). The Evaluator reveals to the 3rd party his raw data value ('0' or '1'), whereupon the 3rd party provides the Evaluator that token (but does not reveal the other token possibility). With OT, Bob receives his token and Alice does not know which token was provided, thus keeping from Alice any knowledge of Bob's value.

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: 1139045

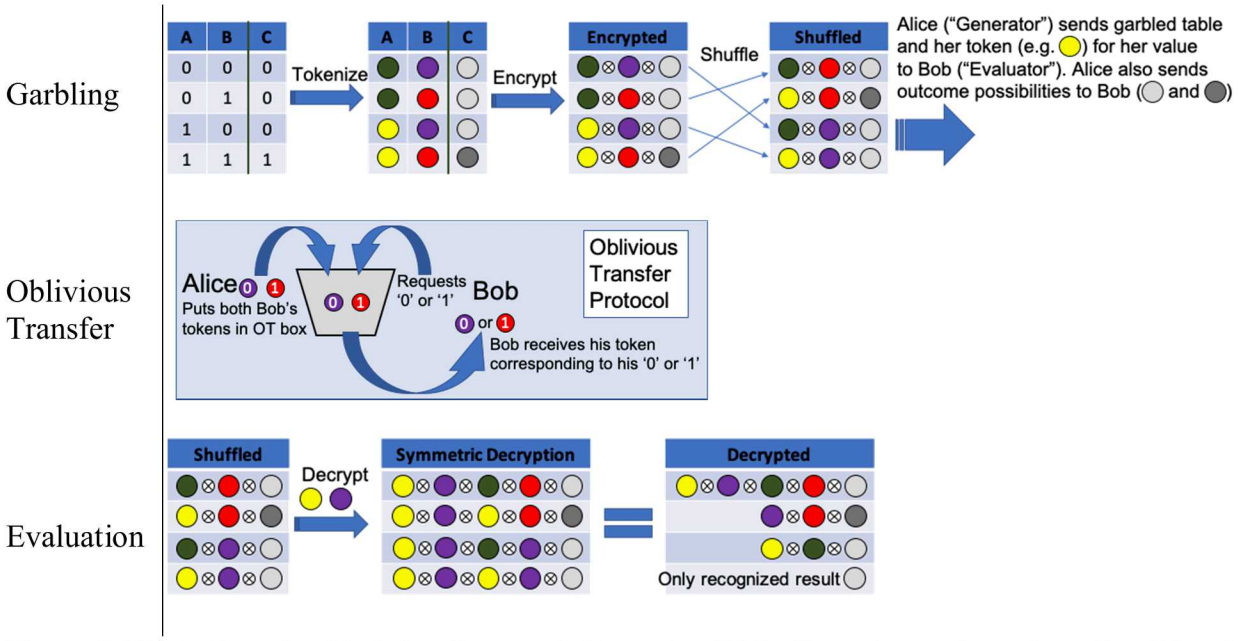


Figure 5. Illustration of garbled circuit protocol. Generator ("Alice") creates random tokens for all input and output possibilities; Oblivious transfer is used to provide the Evaluator ("Bob") with his token; and Evaluator decrypts the table to identify the correct output token.

Now, the Evaluator ("Bob") has all the information he needs to decrypt the garbled gate truth table as shown in the third row of Figure 5. Bob simply decrypts each row of the truth table using his and Alice's tokens, and then sees which of the decrypted tokens matches one of the Column C tokens that Alice gave to him (or else Bob can show the decrypted tokens to Alice and ask her which is the correct one).

Note that the above garbling and evaluation process is done for every gate within the circuit, hence the term "garbled circuit". Another picture of the garbled circuit evaluation process, using an analogy of keys and locks, for the comparator circuit of Figure 4 is shown in Figure 6. One can see how such garbling and evaluating of every gate can quickly increase computation requirements as a circuit becomes large. Efficiency gains are possible, such as using an Oracle for random numbers and re-using one random number for XOR gates [18]. Parallelization is also possible; or doing computations in chunks. One hardware-based acceleration possibility could be to use FPGAs for the computations, for which computing on digital circuits is their main purpose.

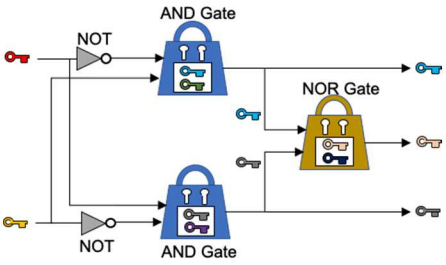


Figure 6. Picture of garbled comparator circuit evaluation process, using keys and locks as an analogy.

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: 1139045

As described in our previous report [21], the garbled circuit protocol we are using is secure for the semi-honest scenario ("passive security"). This means that all parties follow the protocol exactly, including submitting their actual data rather than false data, but will try to uncover the other party's input values. Active security is possible, which could deal with a party not following the protocol, at additional computational cost [22], but still does not account for false data input. There may be complementary methods to authenticate data, such as the SNL-developed EDAS device [23], and we are investigating other cryptographic techniques as well.

We have incorporated the garbled circuit protocol, including Oblivious Transfer, into a Python library we call *CypherCircuit*. For further code details, please see our companion submission authored by Mitch Negus *et al*, "Garbled circuits for enabling privacy preserving safeguards". Two primary focus areas going forward are (1) to increase the speed (efficiency) of our *CypherCircuit* code, possibly including adjusting the protocol currently being used [18]or perhaps re-programming into C/C++ rather than Python; and (2) investigate enhanced security beyond our current passive security paradigm.

CONCLUSIONS

This work is the first 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 have prioritized anomaly detection methods that would be most valuable in extending and improving the capabilities of the existing GS method, and commenced the work on developing, implementing, and testing these new methods. One of our 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 [5]. We have introduced and described a framework by which adoption tradeoffs of DLT for improved CoK are being objectively evaluated. The described resilience framework should be of independent interest, as it can be adapted to inform adoption tradeoffs of other technical proposals, as well as can quantify the resilience of a system (e.g., other aspects of current safeguards practice) in isolation. The two-party formulation of MPC called "garbled circuits" can be used to enable sharing of nuclear data, potentially of safeguards relevance, towards calculating a function result without ever exposing the raw data to the other party. We have developed a Python library called *CypherCircuit* that automatically creates garbled circuits following the protocol of Kolesnikov [18], and are using the Oblivious Transfer protocol to provide the circuit evaluator with his decryption token corresponding to his input value. Herein we have described in pictorial format the circuit garbling and evaluation processes to provide a visual means to understand the protocol. Running of the code can become slow as the circuit size increases, but we are investigating methods to improve speed and efficiencies. The protocol is secure in the semi-honest adversary case (passive security), and we are investigating the possibility of security in the case of a malicious adversary. If we are successful in developing an integrated platform that combines these three core data analysis and management methods, this will be a significant step towards practical applicability of the proposed technologies to authentication, protection, and analysis of the actual safeguards data.

REFERENCES

[1] C. Ramos, C. Pickett, "Defense Nuclear Nonproliferation Research and Development Initiatives in Data Science for Safeguards", ESARDA Symposium, Stresa, Italy, May 14-16, 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: 1139045

[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] E. Keogh, J. Lin, and A. Fu, “Hot SAX: Efficiently finding the most unusual time series subsequence”, Fifth IEEE International Conference on Data Mining (ICDM), pp. 8 (2005).
[7] Y. Zhu, Z. Schall-Zimmerman, N. S. Senobari, C.-C. M. Yeh, G. Funning, A. Mueen, P. Brisk, and E. J. Keogh, “Matrix Profile II: Exploiting a Novel Algorithm and GPUs to Break the One Hundred Million Barrier for Time Series Motifs and Joins”, IEEE 16th International Conference on Data Mining (ICDM) (2016), pp. 739–748 (2016).
[8] D. Blair and N. Rowe, "Global Perspective on Continuity of Knowledge: Concepts and Challenges", Institute of Nuclear Materials Management Annual Meeting, 2014.
[9] S. Frazar, C. Joslyn, R. Singh and A. Sayre, "Evaluating Safeguards Use Cases for Blockchain Applications", PNNL-28050, Oct 2018
[10] C. Vestergaard, "Better Than a Floppy, The Potential of Distributed Ledger Technology for Nuclear Safeguards Information Management", Stimson Center Policy Brief, Oct 2018.
[11] M. Westerkamp, F. Victor and A. Küpper, "Blockchain-Based Supply Chain Traceability: Token Recipes Model Manufacturing Processes", 2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), Halifax, NS, Canada, 2018.
[12] C. T. Hu, D. R. Kuhn and D. F. Ferraiolo, "Access Control for Emerging Distributed Systems", IEEE Computer, 2018.
[13] I. A. E. Agency, "Safeguards Implementation Practices Guide on Provision of Information to the IAEA", IAEA Servies Series No. 33, Vienna, 2016.
[14] B. Meppen, T. Bean, R. Haga, K. Moedl, J. Sanders and M. A. Thom, "Validation of Nuclear Material Control and Accountability (MC&A) System Effectiveness Tool (MSET) at Idaho National Laboratory”, Institute of Nuclear Materials Management Annual Meeting, 2008.
[15] G. D. Wyss, J. L. Darby, P. G. Dawson, K. J. Page and E. E. Ryder, Risk-Based Decision Approaches for Safeguards & Security Management, Albuquerque, NM: American Nuclear Society Winter Meeting, 2006.
[16] M. Turnquist and E. Vugrin, "Design for resilience in infrastructure distribution networks", Environment Systems & Decisions, vol. 33, no. 1, pp. 104-120, 2013.
[17] M. Galiardi, A. Gonzales, J. Thorpe, E. Vugrin, R. Fasano and C. Lamb, "Cyber Resilience Analysis of SCADA Systems in Nuclear Power Plants", 28th Conference on Nuclear Engineering & Joint With the ASME 2020 Power Conference, Anaheim, CA, 2020.
[18] V. Kolesnikov and T. Schneider, "Improved garbled circuit: Free XOR gates and applications", in International Colloquium on Automata, Languages and Programming (ICALP 2008), Berlin, 2008.
[19] M. Naor and B. Pinkas, "Efficient oblivious transfer protocols”, Proceedings of the twelfth annual ACM-SIAM symposium on Discrete algorithms, Philadelphia, 2001.
[20] M. Bellare and S. Micali, "Non-interactive oblivious transfer and applications", in Advances in Cryptology - CRYPTO '89 Proceedings, New York, 1990.
[21] A. Solodov, D. Farley, C. Brif, N. Pattengale, Y. Gao, J. Lin, M. Negus and R. Slaybaugh, "Development of Novel Approaches to Anomaly Detection and Surety for Safeguards Data”, 60th Annual Meeting of the Institute for Nuclear Materials Management, Palm Desert, 2019.
[22] C. Baum, I. Damgard and C. Orlandi, "Publicly Auditable Secure Multi-Party Computation", International Conference on Security and Cryptography for Networks, Cham, 2014.
[23] M. Thomas, G. Baldwin, J. G. Goncalves, A. Smejkal, R. Hymel, R. Linnebach, L. Deschamp, S. Johnson and M. Rue, "Testing the Enhanced Data Authentication System (EDAS)”, IAEA Symposium on International Safeguards, Vienna, 2014.

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: 1139045