SAND2024-07310

SANDIA REPORT
SAND20XX-XXXX
Printed Click to enter a date

Integrity Enhancing Protocols (IEP) Evaluation Framework

Romuald Valme, Adam Beauchaine, Minami Tanaka, Christopher Lamb

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico
87185 and Livermore,
California 94550

Sandia
National
Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by National Technology & Engineering Solutions of Sandia, LLC.

NOTICE: This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, make any warranty, express or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represent that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government, any agency thereof, or any of their contractors or subcontractors. The views and opinions expressed herein do not necessarily state or reflect those of the United States Government, any agency thereof, or any of their contractors.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone:	(865) 576-8401
Facsimile:	(865) 576-5728
E-Mail:	reports@osti.gov
Online ordering:	http://www.osti.gov/scitech

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Rd
Alexandria, VA 22312

Telephone:	(800) 553-6847
Facsimile:	(703) 605-6900
E-Mail:	orders@ntis.gov
Online order:	https://classic.ntis.gov/help/order-methods/

ABSTRACT

Protocols play an essential role in Advance Reactor systems. A diverse set of protocols are available to these reactors. Advanced Reactors benefit from technologies that can minimize their resource utilization and costs. Evaluation frameworks are often used when assessing protocols and processes related to cryptographic security systems. The following report discusses the various characteristics associated with these protocol evaluation frameworks, and derives a novel evaluative framework tailored to operational technology (OT) systems. Section 2 provides a literature review of related works. Various attributes are introduced to create a pool of potential categories and metrics to consider. Section 3 selects and describes the categories for the novel framework. This section also presents Minimega, which is the testbed environment that will be used for producing the testbed. Minimega can host and orchestrate virtual machines, it can also network them creating custom topologies. Section 4 explains and justifies selected testbed attributes. It also defines each attribute in the chosen categories. This section describes how these attributes are measured and their importance to this framework and how they will be used to evaluate various integrity enhancing communication protocols.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

This page left blank

EXECUTIVE SUMMARY

Protocols are ubiquitous within Advanced Reactor systems. A powerful feature of electronic devices is their ability to communicate with one another. The ability to transfer data from one machine to another allows for the engagement in countless applications. In advanced reactors the authenticity, integrity and confidentiality of data must be considered. Through competent evaluation costs of operating a facility can be minimized and usage of essential resources in memory and processing can be reduced. This report will be evaluating integrity protocols, which can verify that data has not been altered or tampered with and often even the identity of the sender.

This work contains a review of relevant frameworks for evaluating cryptographic integrity-providing protocols. Section two presents the capabilities and characteristics of various evaluation frameworks found in prior work. Evaluation frameworks focus on measuring specific metrics to determine the performance of a particular protocol. This section identifies attributes and begins to form categories of metrics that have been observed in previous work. These categories span a range of potential attributes, which allows for a thorough selection process for each category. Out of the various attributes hardware usage and running time statistics during critical processes appear to be popular and data rich metrics to capture.

Section three starts the attribute selection process. This section explores the evaluation attributes mentioned in the literature review of related work. Four categories of attributes are selected and described. These categories are protocol running time, endpoint storage requirement, protocol hardware performance, and existing security evaluation level (cryptographic hardness).

When running any cryptographic procedure, there is a precise execution time. The protocol running or execution time records the amount of time necessary to perform cryptographic digital signature functions. The latency created by an algorithm is a critical parameter. Once recorded one can compare the average latency of a protocol to that of others for the best performing process. These latency measurements are coupled with theoretical analysis of protocol worst case performance.

Endpoint storage and key management requirements create an understanding of where the keys of a cryptographic procedure will be located and what kinds of protections and restrictions are in place. Protocols typically have their own encryption, decryption, and storage practices. For digital signatures, the processes would be the signing and verifying of the signature, which may require keys to be used as a proxy for identifying a user or system. This method implies secure keying material creation and storage. The storage requirements of the keys in these procedures must be evaluated, in addition to the storage requirements of cryptosystem components themselves.

Protocol hardware performance consists of measuring Central Processing Unit (CPU) utilization and memory usage for protocol operations. The efficiency of an algorithm is highly coupled to the amount of CPU and memory it uses. Typically, higher usage correlates to poorer performance. This work plans to use standards such as the NIST Cybersecurity Evaluation Framework which can be used to grade algorithms on their ability to detect and protect against threats. This guidance can serve as a gauge of cryptographic hardness and attack resistance provided by a protocol. These levels can be used to assess performance in the topologies and scenarios protocols will encounter.

This section also explains testbed and environment implementation details. An evaluation of a protocol relies on the underlying infrastructure in which the protocol operates. To realistically test a protocol, a researcher must utilize a distributed network testbed which can emulate typical traffic. Minimega is a tool that allows its user to emulate virtual networks. With it, entire virtual machines as well as their interconnections can be emulated. To accomplish the evaluation of integrity protocols, the use of this virtualized testbed to emulate typical communication is necessary. Through Minimega, thorough quantitative analysis of several categories of algorithms spanning various possible environments becomes possible.

Section four introduces each attribute selected for a category. This section describes how each attribute is captured in the emulation environment and discusses why it has been selected. Running time attributes are captured by placing a timer around the programmed cryptographic functions. Storage requirements are determined by calculating the available space on a device. The access rights and mechanisms to secure keys must also be ascertained. Hardware attributes are observed by utilizing standard linux commands and tools such as process status (ps). These tools can be accessed through Miniweb or the various scripting tools in Minimega. These metrics have been selected because of their utility in deciding if performance requirements are met in resource-constrained OT systems. The NIST classification levels provides general security quantification, which can be applied to a wide range of OT environments. It does this through the comparison of a protocol against others such as AES or SHA.

ACRONYMS AND TERMS

Acronym/Term	Definition
IoT	Internet of Things
OT	Operational Technology
MAC	Message Authentication Code
RAM	Random Access Memory
NIST	National Institute of Standards and Technology
RSA	Rivest–Shamir–Adleman (Public Key Cryptosystem)
FIPS	Federal Information Processing Standards
PQC	Post Quantum Cryptography
CPU	Central Processing Unit
KVM	Kernel-based Virtual Machine
QEMU	Quick Emulator
VM	Virtual Machine
VCPU	Virtual Central Processing Unity
AES	Advanced Encryption Standard
SHA	Secure Hashing Algorithm

1. INTRODUCTION

Evaluation frameworks are used to create structure for the testing of some process. These frameworks serve as roadmaps, guiding evaluators through the process of gathering, analyzing, and interpreting data to make informed decisions. From healthcare to engineering, many industries have processes that are regularly evaluated. Through this structure researchers can compose clearly repeatable experiments and observe results in a formal manner. This work constructs an evaluation framework for the assessment of integrity enhancing protocols in OT systems. Experiments on these protocols will be executed in a virtualized networked environment.

These frameworks consist of attributes which perform some function necessary for the evaluation to take place. Attributes are typically in one of three categories: definition/configuration, execution, and analysis [1]. This review includes certain modular attributes, detailed in a subsequent section. Many researchers have developed their own unique frameworks, which contain some shared and unique attributes. Specific details of an attribute may include operational metrics. The capture and analysis of these metrics will be a crucial part of this framework and will provide readers with insight on the most efficient and worthwhile protocols for OT systems.

The evaluation framework falls into a process flow which guides experimentation. Each step of the flow is critical in obtaining results. In an evaluation framework, the inputs, environment, activities, outputs, and outcomes should be clearly understood, as shown in Figure 1. For these experiments the inputs to the framework will be the selected protocol. The activities will be the cryptographic procedures selected. The environment will leverage a virtualized network testbed. The output will be the recorded metrics.

This report describes the frameworks currently in use and what their strengths and weaknesses are. The framework development section discusses what attributes should be considered in this case, where we are looking specifically at the integrity enhancing properties of communication protocols. It also covers what the ideal environment for testing would be. The analysis and evaluation section goes deeper into the metrics selected and rationale.

Advanced reactors benefit from the evaluation of protocols through a robust framework. Effective evaluation requires careful planning, data collection, analysis, and interpretation. By employing rigorous evaluation organizations can improve accountability, cost, learning, and decision-making to ultimately enhance the effectiveness and impact of their initiatives.

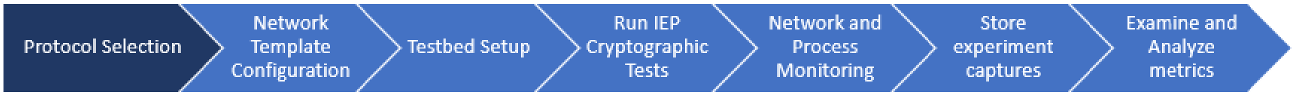
Protocol Selection
Network Template Configuration
Testbed Setup
Run IEP Cryptographic Tests
Network and Process Monitoring
Store experiment captures
Examine and Analyze metrics

Figure 1: Process Flow for Experimentation.

2. RELATED WORK

This section summarizes prior work involving evaluation of a wide range of cryptographic schemes, as well as previous efforts specific to integrity-guaranteeing protocols. It examines previous evaluations of integrity-guaranteeing protocol performance, as well as existing material for protocol security labeling. The fusion of these insights allows for the observation and documentation of specific framework attributes.

Cryptographic evaluation schemes are a commonly used tool for determining protocol suitability across a range of information environments. [2], [3], [4], [5], [6] Among environments analyzed, environments with resource constrained devices, such as the Internet of Things (IoT), are researched more thoroughly due to the increased importance of protocol performance on endpoint devices with very limited computational resources. There is additionally a larger, consumer-focused market for these kinds of devices. Pereira et al. [7] leverage time and energy consumption metrics within a performance evaluating framework for the Intel Edison IoT platform. Fotovvat et al. [3] instead specify their efforts toward endpoint hardware, using a Raspberry Pi device to simulate and record the hardware usage statistics. The recorded metrics are power consumption, random access memory (RAM) usage, and overall execution time, recorded on various block and permutation-based cryptographic cyphers. In contrast with previous efforts, the proposed evaluative framework will target endpoint devices specifically used in operational technology (OT) environments. The proposed framework contrasts with existing approaches by focusing on specific evaluation of integrity-guaranteeing protocols for these systems.

Integrity has been established traditionally in information systems through the usage of several common approaches. The use of public key cryptographic systems for digital signatures has seen near ubiquitous usage for message integrity verification in distributed systems due to its advantages over previous existing approaches such as message authentication codes (MAC). The usage of Galois Counter mode for block ciphers allows for integrity management of large amounts of networked data. With respect to the former approach, Hossain et al. [8] evaluate public key cryptosystems in the context of large-scale data encryption against symmetric ciphers such as AES, noting the overall transaction times of data transfers. A similar performance metric is used in the work of Jansma et al. [9] in which RSA and Elliptic Curve digital signatures are compared with respect to key generation and signing times. The approach [9]□□ , but considers additional performative metrics beyond execution time for cryptographic schemes. This allow framework to better identify

Researchers have sought to quantify security levels of various cryptographic algorithms through theoretical principles as well as security frameworks. Boreale et al. [10] develop a finite set of symbolic semantics for protocol execution, allowing for security evaluation of authentication and secrecy protocols. This symbolic approach allows for the security quantification of more complex systems, including those involving multiple protocols. An alternative evaluation framework, such as the *NIST FIPS 140-3 Security Requirements for Cryptographic Modules,* relies less on constructed formalisms and direct analysis of protocol complexity. This work leverages similarly straightforward existing analytical frameworks for the purposes of quantifying security as a single component of overall protocol effectiveness within an OT system. The usage of pre-existing frameworks allows for the simplification of this framework, and the lack of reliance on constructed formalisms allow for a clearer immediate understanding of overall protocol security performance.

Due to the strong similarities, and often identical hardware requirements, this framework may leverage the insights gained from IoT evaluative frameworks, as well as more general case research, for the evaluation of integrity-guaranteeing protocols in OT systems. This work further identifies existing protocol performance metrics of high relevance to OT architecture and contrasts the proposed framework with these previous efforts.

3. FRAMEWORK DEVELOPMENT

Despite the wide range of existing work regarding security and performance evaluation on information systems, analysis specific to integrity guaranteeing protocols in OT systems is much sparser. These environments have more specific performance and security needs than traditional IT and IoT systems, and OT system designers are often forced to rely upon security analysis on systems that may only feature a handful of similarities to their own. This section reviews existing security protocol evaluation attributes, as well as previously developed frameworks for similar purposes. Additionally, this section develops topological design and implementation goals for a testbed to quantify the metrics discussed.

3.1. Framework Attribute Selection

Prior work has identified many quantitative attributes of cryptographic protocols for security and performance evaluation. In one of the most thorough frameworks developed for performance, Deshpande et al. [11] identify information throughput, system power consumption, system gate area, and transaction latency as areas of importance for block cipher performance evaluation. Due to the narrowed focus of this work on integrity guaranteeing protocols within OT systems, many of these attributes, such as throughput and latency, are not as important. The body of work specific to integrity-guaranteeing protocol evaluation is considerably less developed than other paradigms. As a result of this, this work leverages such existing work, as well as general knowledge of protocol evaluation, in developing framework attributes.

Certain performance metrics prove to be uniquely valuable in the evaluation of OT and IoT communication systems. Running times of key generation, message encapsulation, and de-encapsulation have shown to be historically valuable in protocol evaluation, [12] and are of equal importance to resource-constrained OT systems. Key management hierarchies [13] have been shown to be an effective manner of visualizing key access and storage requirements in distributed systems. Such techniques have been leveraged in combination with analysis of long-term key storage requirements, which other researchers have noted as being highly variable between protocols. [14] Other work [15] has noted the importance of small key values as a necessary feature for widespread adoption. Additional storage requirements, such as cryptosystem storage requirements, or hyperparameter value storage in the case of neural algorithms, must also be accounted for. Short term memory performance is additionally an area of concern and may even cause device failure in certain scenarios. [14] As with running times, memory usage may be subdivided into key generation and communication service requirements. Finally, the NIST security level of a given protocol, including the module security levels FIPS 140-2 [16] and NIST PQC Security Levels [17] have seen relatively common usage for simple quantitative security evaluation.

As a result of this analysis, this work has identified a set of metrics, which may be organized into four broader categories:

- **Protocol Running Time:** This component of analysis includes real time, as well as asymptotic analysis for each protocol selected. Running time metrics for digital signature schemes may be further subdivided into key generation and inter device communication and are typically recorded in microseconds.
- **Endpoint Storage Requirements:** This component identifies both key and cryptosystem storage requirements on an endpoint device. Due to variable key lengths between protocols,

coupled with often low storage capacities of OT environments, this is a significant metric in measuring the difficulty of practical implementation.

- **Protocol Hardware Performance:** This work leverages standard analytical tools to explore CPU and memory usage for protocol components. Due to the time-critical nature of OT systems, additionally investigate latency overheads or levels of system jitter associated with individual protocols.
- **Existing NIST Security Evaluation Level:** This work includes the relevant NIST security classification levels for each protocol tested, this serves as a relative indicator of overall protocol attack resistance and cryptographic hardness. When coupled with findings on performance, this component offers additional insight into overall protocol usefulness to OT security administrators. While these scores are not provided for all evaluated protocols, NIST classification parameters allows them to be easily derived based on protocol cryptographic hardness. We detail this further in Section 4.

3.2. Existing Evaluative Frameworks

Existing evaluative frameworks have focused on individual aspects of systems and protocols. Based on research goals, these frameworks may be qualitative [18] [19] or quantitative, [20] [21] with respect to measured attributes. Bonneau et al. [18] leverage the former, qualitative approach to evaluate a significant amount of system design approaches for user password replacement, significantly reducing the required amount of data gathering to evaluate such protocols. Lafourcade et al. [20] take the latter, quantitative approach in the analysis of protocols based on the arithmetic method of modular exponentiation, detailing significant hardware statistics for each documented approach, and comparing to the respective baseline of Diffie-Hellman. Both paradigms have seen widespread usage in prior evaluative efforts, this work's approach identifies strengths and weaknesses of these paradigms and existing frameworks, to leverage in this proposed framework.

The usage of either of these paradigms most commonly relates to the number of candidates to be analyzed. Large amounts of candidates are more likely to leverage a simpler, qualitative approach, for streamline data gathering across a high number of required experiments. Qualitative approaches may be significantly more involved and increase the required data gathering significantly. However, they can additionally present a more detailed, granular view of candidate performance and suitability for a given task.

3.3. Testbed Evaluation

This section outlines the ideal testbed features and topological design to sufficiently observe protocols functioning within an OT system. Based on the protocols selected, custom scenarios will be devised to observe behavior and produce attribute metrics for the evaluative framework. Due to the focus on OT systems, these experimental scenarios will be implemented in an environment aimed at simulating OT endpoint devices. These devices feature specific networking practices for deployment in industrial systems, as shown in Figure 2. To preserve modularity and usability, the recommended testbed should be virtual, with a focus on standard networking protocols to allow for ease of configuration, as well as experimental data capture. The team intends to use the Sandia developed tool Minimega [22] for this purpose. Minimega has rapid deployment capabilities, low configurational complexity, and integrated data capturing tools leading to easier metric collection. Minimega also supports inspection and capture of network and file system activities on all virtualized endpoint machines, simplifying performance measurement.

Minimega interacts directly with the KVM hypervisor and QEMU [23] emulation platform, allowing for compatibility with most Debian-based Linux systems. Minimega requires no external software stack or complex initial configuration, allowing for fast, efficient deployment of a wide array of experimental testbed scenarios. OpenVSwitch is leveraged for an internal switching stack, and VMbetter allows for the export of virtual hosts into many common disk image formats. Protonuke, a layer 3 traffic generation module, allows for diverse network conditions for experimentation. These features allow for the ease of recording and exporting scientific results, which in turn allows for the replicability of protocol performance evaluation.

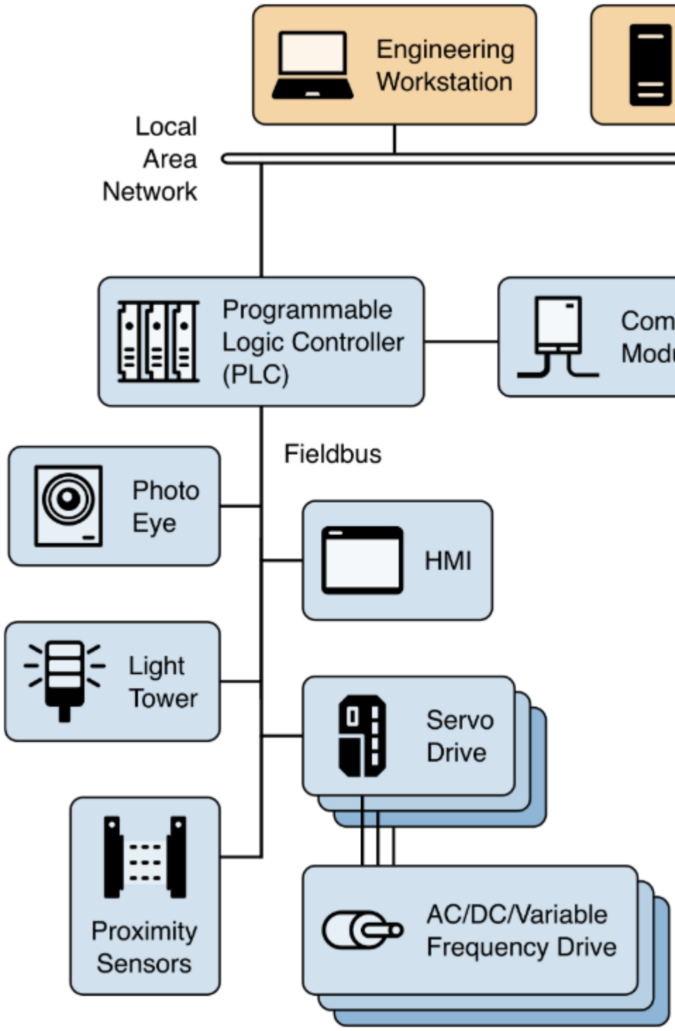

Figure 2: A typical OT topology using fieldbus in tandem with standard networking protocols. [24]

4. FRAMEWORK ANALYSIS AND EVALUATION

To leverage this testing design within the evaluative framework, metrics are selected with a focus on quantifiability within a virtualized testbed environment. Minimega allows for extensive data gathering within configured environments, including VM disk, memory, and VCPU usage statistics. Given the lack of formalization in this space, quantifiable metrics should provide system engineers with more relevant data to meet performance and security goals relative to their environments. This framework includes external security quantification through the usage of NIST Protocol Security Classifications for relevant protocols. This framework additionally includes asymptotic analysis for key generation on relevant protocols as a second externally derived metric. A complete list of these metrics and groups is provided below in table 2. Included in the table is the metric of measurement itself, as well as the relevant method for attaining these measurements in testing. While this work refrains from specific metric weights, it emphasizes running time and performance metrics as the primary focus of testing, due to their high relevance to OT system development.

This work derives the attribute groupings and measurement strategies from the prior section detailing Framework Attribute Selection. Attribute selection was additionally based on the capacity to measure these components as part of a Minimega experimental workflow. Users may instrument collection of these features during a single protocol evaluative experiment through Minimega. These values may be observed and recorded after a workflow, as well as retrieved from algorithm documentation. The lack of weights in the model allows for the highlighting of various performance attributes of protocols, this is useful because of the wide variety of use cases and improvement targets various OT systems may seek to implement.

Attribute Group	Attribute	Metric	Notes
Protocol Running time	Key Generation Running Time	Microseconds (µs)	Measured in function call
	Key Generation Asymptotic Worst Case	Big O Notation (O(n))	Derived theoretically
	Key Encapsulation Running Time	Microseconds (µs)	Measured in function call
	Key Decapsulation Running Time	Microseconds (µs)	Measured in function call
Endpoint Storage Requirements	Key Storage Requirements	Bytes	Single key storage
	Cryptosystem Storage Requirements	Bytes	Algorithm component storage requirements
Protocol Hardware Performance	Cryptosystem Average CPU Utilization	Percentage	Average usage over a single transaction
	Cryptosystem Average Memory Usage	Bytes	Average usage over a single transaction
	Network Transmission Time	Milliseconds (ms)	Measured in function call
Existing Security Evaluation Level	NIST Security Classification of Modules (FIPS 140-2 [16])	Integer Scale (1-4)	Physical module security
	NIST PQC Security Project Levels [17]	Integer Scale (1-5)	Protocol security in relation to classical protocols

Table 1: The Framework's Analyzed Metrics and Groups

Within the category of protocol running time attributes, the selected framework attributes concern the generation and communication components of evaluated cryptosystems. Measurements are conducted in units of time in this section, due to the importance of timing specific system components to meet potential system requirements. The system provides timing measurements of key generation, as well as "encapsulation" and "decapsulation" of signed message data. While the usage of encapsulation and decapsulation traditionally refers to a symmetric key exchange algorithm such as Diffie-Hellman, this framework assumes these operations are conducted on a message of any type. These timing metrics may be evaluated in a simple testbed environment through simple hook functions applied to standard cryptographic program libraries. These enable the framework to provide microsecond measurements of each individually called component. The one distinct attribute in the category is the asymptotic worst case running time of key generation. We limit theoretical analysis to the key generation, due to this step requiring the most algorithmic complexity, and include these measurements in the standard Big O Notation.

The endpoint storage requirement category addresses storage requirements for evaluated cryptosystems. This is distinct from runtime memory (RAM) requirements for a cryptosystem, as these metrics specifically refer to disk storage requirements of evaluated protocols. The lack of storage on OT devices has been noted as a concern in evaluations and implementations of cryptographic schemes in previous work. The framework attributes in this section concern the storage requirements of keys as well as cryptosystem components themselves. Key storage is self-explanatory, cryptosystem storage requirements refer to any library or computational storage components required for system operation, such as locally hosted neural network components for neural cryptographic schemes. Storage requirements are calculated through simple file system measurements to reveal how much space a given component requires. This may additionally be observed in library or cryptosystem documentation.

Protocol hardware performance is distinct from running time in that the framework conducts direct observations on hardware in use metrics. These metrics are similarly useful in determining if performance requirements are met in resource-constrained OT systems. Measurements are conducted by analyzing hypervisor control tools such as Miniweb [22] during experimentation. These tools may be orchestrated to record this information automatically during an experimental run of the system. The framework evaluates these results through comparison with existing system hardware evaluations, such as the work of Beg et al. [25]

Existing security evaluation levels are the final component of the evaluative framework and offer a system independent manner of quantifying security performance across existing implementations. The benefits of such an evaluation enable system administrators to better conceptualize the security benefits of specific protocols when considering organizational implementation. Most analysis of integrity providing systems in prior work has been qualitative as opposed to quantitative. However, several approaches based on formalism have gained traction for integrity quantification on a system wide level. [26] Given the diversity of approaches this framework seeks to evaluate, the ability to quantify relative cryptographic hardness, or the "amount" of integrity provided is critical. When applicable, the framework includes the NIST Security Classification of Module (FIPS 140-2 [16]) and PQC Security Levels. [17] The inclusion of these two metrics allows the framework to provide quantification in the absence of system or program-specific formalisms. This approach was selected due to its general-case applicability to a wide range of OT systems. While program or domain specific evaluations will provide more detailed analysis in most cases, this approach presents general integrity performance of protocols and cryptographic hardware systems.

NIST FIPS 140-2 [16] provides an integer scale for hardware security of cryptographic modules. This scale runs from 1-4 and is based on physical security and evaluates hardware capacity to respond to tampering and other physical attacks. The testing environment, while virtual, may simulate the modules that would support these protocols in OT environments. By including corresponding module security levels in a realistic industrial setting, a framework user may quantify hardware module security as a component of this framework. For the security of protocols themselves, this framework leverages and expands the NIST PQC project security classification levels. This integer scale runs from 1-5 and provides post quantum protocol specific security quantification based on cryptographic hardness in comparison with classical approaches. These levels are detailed below:

- **Level 1** – At least as hard to break as AES-128
- **Level 2** – At least as hard to break as SHA-256
- **Level 3** – At least as hard to break as AES-192
- **Level 4** – At least as hard to break as SHA-384
- **Level 5** – At least as hard to break as AES-256

By leveraging fastest known methods of breaking these protocols, a user may directly quantify which level a protocol belongs to. For example, CRYSTALS-Dilithium public key cryptosystem scores a 5 on the scale due to its usage of a high dimensional lattice problem being mathematically comparable with AES-256 (14 rounds). These comparisons and measurements may be applied to additional protocol schemes, even those not yet included in the PQC project list and allows for a user to quantify a wide variety of diverse protocols for integrity provision.

Given the resource constrained nature of OT systems, this work draws several parallels to evaluative efforts in the IoT space. These include the analysis of system storage requirements, due to the high variability of storage requirements between cryptosystems [14]. Additionally, CPU, memory, and network transmission time are all high relevance to integrity-guaranteeing cryptosystems for data in transit. To meet usability goals in relation to this testbed, measurements such as gate complexity and electricity usage metrics are omitted. Despite being detailed in existing work, the challenges of measuring and reporting on these metrics in a fully virtualized environment is beyond the scope of this work. Likely the most important metric in this framework is that of real time measurements on protocol running times. Such data will enable conclusions regarding protocol viability in OT systems, given the potential consequences of jitter and delays in OT environments.

Performance data will be organized into numerical values for each category and attribute. It will not be combined into a composite score. This is done due to the diversity of metrics evaluated. This diversity means that system designers will have to observe individual framework components to derive valuable insights. Even per-category composite scores may obscure meaningful results, such as running time or storage requirements, given the wide range of protocol algorithmic complexities and storage models. This lack of result abstraction allows for framework users to derive optimal execution contexts for selected algorithms and replicate the corresponding scenarios for their own adoption.

5. CONCLUSION

Evaluation frameworks are essential tools for determining the effectiveness of processes. This tool can be used to examine and grade integrity enhancing protocols. This report describes a modern framework. The framework prioritizes attributes which have been shown to be valuable by related efforts throughout industry. Section two describes the various works. These efforts incorporate attributes such as key generation running time and CPU utilization which are assessed for their relevance to the team's goal of evaluating integrity enhancing protocols. Section three introduces four relevant attributes categories. These are Protocol Running Time, Endpoint Storage, Protocol Hardware Performance, and Existing Security Evaluation Level. These categories were chosen because of their ability to assess the performance of protocols in constrained OT environments. Section three also introduces an experimental environment, which can generate high-fidelity quantitative results. This section discusses its ability to simulate OT devices, allowing researchers to tailor their efforts towards OT environments. Section four delves into what attributes in each category we intend to use and describes how each is recorded. Why they are included in this framework is also discussed.

This framework is embedded in an evaluation process which instructs researchers on how to perform the experiment. This process starts with protocol selection, and then moves onto setting up the environment. After setting up the environment, testing on the protocol is performed, which is then followed by the analysis of the generated metrics.

Future work could include incorporating Hardware-in-the-loop systems. These devices can be utilized to extract more device characteristics and metrics. By guiding evaluators through the process of data collection, analysis, and interpretation, evaluation frameworks enhance accountability, learning, and decision-making processes within organizations. Ultimately the use of rigorous evaluation frameworks enables stakeholders to make informed decisions, optimize resource allocation, and maximize the impact of their efforts in addressing complex challenges and achieving desired outcomes.

6. REFERENCES

[1] J. Friginal and D. Andres, "REFRAHN: A Resilience Evaluation Framework for Ad Hoc Routing Protocols," *Science Direct,* vol. 82, pp. 114-134, 2015.

[2] P. S. Kedara Deshpande, "Performance Evaluation of Cryptographic Ciphers on IoT Devices," *arXiv,* 2018.

[3] A. Fotovvat, G. M. E. Rahman, S. S. Vedaei and K. A. Wahid, "Comparative Performance Analysis of Lightweight Cryptography Algorithms for IoT Sensor Nodes," *IEEE IoT Journal,* 2020.

[4] C. B. U. Ç. &. S. K. Pejman Panahi, "Performance Evaluation of Lightweight Encryption Algorithms for IoT-Based Applications," *Arabian Journal for Science and Engineering,* 2021.

[5] S. Maitra, D. Richards, A. Abdelgawad and K. Yelamarthi, "Performance Evaluation of IoT Encryption Algorithms: Memory, Timing, and Energy," *SAS,* 2019.

[6] V. A. Thakor, M. A. Razzaque and M. R. A. Khandaker, "Lightweight Cryptography Algorithms for Resource-Constrained IoT Devices: A Review, Comparison and Research Opportunities," *IEEE IoT Journal,* 2021.

[7] G. C. C. F. Pereira, R. C, A. Alve, F. L. d. Silv and Roberto M. Azevedo, "Performance Evaluation of Cryptographic Algorithms over," *SPEWN,* 2017.

[8] A. Hossain, B. Hossain, S. Uddin and S. Imtiaz, "Performance Analysis of Different Cryptography Algorithms," *International Journal of Advanced Research in Computer Science and Software Engineering,* 2016.

[9] N. Jansma and B. Arrendondo, "Performance Comparison of Elliptic Curve and RSA Digital Signatures," *UMICH Internal,* 2004.

[10] M. Boreale and M. G. Buscemi, "A Framework for the Analysis of Security Protocols," *International Conference on Concurrency Theory,* 2002.

[11] P. Singh and K. Deshpande, "Performance Evaluation of Cryptographic Ciphers on IoT Devices," *ICTRCET,* 2018.

[12] F. Lauterbach, P. Burdiak, F. Richter and M. Voznak, "Performance Analysis of Post-Quantum Algorithms," *Telecommunications Forum,* no. 9, 2021.

[13] N. Kumar and A. Mathuria, "Comprehensive Evaluation of Key Management Hierarchies for Outsourced Data," *Cybersecurity,* vol. 2, no. 1, 2019.

[14] A. Fournaris, G. Tasopoulos, M. Brohet and F. Regazzoni, "Running Longer to Slim Down: Post Quantum Cryptography on Memory Constrained Devices," *IEEE International Conference on Omni-layer Intelligent Systems,* 2023.

[15] M. d. F. Dornelles, P. C. d. S. Lara and F. d. R. Henriques, "Performance Evaluation and Comparison of Default and Small Private Key Rainbow Digital Signature Scheme for IoT Devices," *Brazillian Symposium,* 2019.

[16] "Security Requirements for Cryptographic Modules," National Institute of Standards and Technology, 22 March 2019. [Online]. Available: https://csrc.nist.gov/pubs/fips/140-3/final. [Accessed 29 2 2024].

[17] "Post Quantum Cryptography," National Institute of Standards & Technology, 3 January 2017. [Online]. Available: https://csrc.nist.gov/projects/post-quantum-cryptography/post-quantum-

cryptography-standardization/evaluation-criteria/security-(evaluation-criteria). [Accessed 29 February 2024].

[18] J. Bonneau, C. Herley, P. C. v. Oorschot and F. Stajano, "The Quest to Replace Passwords: A Framework for Comparative Evaluation of Web Authentication Schemes," *IEEE Symposium on Security and Privacy,* 2012.

[19] R. Alvarez, C. Caballero-Gil, J. Santonja and A. Zamora, "Algorithms for Lightweight Key Exchange," *UCAml,* 2017.

[20] P. Lafourcade and M. Puys, "Performance Evaluation of Cryptographic Verification Tools Dealing with Algebraic Properties," *LNSC,* 2016.

[21] S. Hallgren, A. Smith and F. Song, "Classical Cryptographic Protocols in a Quantum World," *International Journal of Quantum Information,* 2011.

[22] J. Crussell, J. Erickson, D. Fritz and J. Floren, "minimega v3.0," Sandia National Labs, Albuquerque, NM, 2015.

[23] N.A., "QEMU: A Generic and Open Source Machine Emulator and Virtualizer," QEMU, 4 March 2024. [Online]. Available: https://www.qemu.org/. [Accessed 5 March 2024].

[24] K. Stouffer, M. Pease, C. Tang, T. Zimmerman, V. Pillitteri, S. Lightman, A. Hahn, S. Saravia, A. Sherule and M. Thompson, "NIST Special Publication NIST SP 800-82r3 Guide to Operational Technology (OT) Security," NIST, 2023.

[25] A. Beg, T. Al=Kharobi and A. Al-Nasser, "Performance Evaluation and Review of Lightweight Cryptography in an Internet-of-Things Environment," *ICCAIS,* 2019.

[26] M. R. Clarkson and F. B. Schneider, "Quantification of Integrity," *Math. Struct. In Comp. Science,* 2012.

[27] N. I. o. S. a. Technology, "NIST.SP.800-82r3 Special Report," 2023.

DISTRIBUTION

Email—Internal

Name	Org.	Sandia Email Address
Lon Dawson	8851	ladawso@sandia.gov
Ben Cipiti	8845	bbcipit@sandia.gov
Technical Library	1911	sanddocs@sandia.gov

Email—External

Name	Company Email Address	Company Name
Katya Le Blanc	katya.leblanc@inl.gov	Idaho National Laboratory

This page left blank