Solution Brief

Intel Security
Gramine

Gramine: Protecting Unmodified Linux Applications with Confidential Computing

A Linux-compatible library OS for multi-process applications

Security: A universal requirement

The need for end-to-end data protection has never been greater. As computing moves from on-premises to the public cloud and to the edge, individuals, companies, and organizations are becoming increasingly vulnerable to hacks, data breaches, and malicious attacks. Threats to data have increased, and securing data has become a number one business imperative.

In addition, data owners must deal with a new class of legal requirements as regulations such as GDPR, Schrems II, and HIPAA take effect across the globe, even as breaches of personal data still occur on an all-too-regular basis. At the same time, cloud service providers want to be outside the client "trust boundary" so that even as system administrators they have no visibility into private data (and hence no exposure to subpoenas or other regulatory actions). In today's information economy, organizations achieve data protection by using encryption technologies for data at rest and data in transit, but data still needs to be brought into clear in memory while in use, resulting in a data protection gap.

Rising to the challenges

To meet these challenges, Confidential Computing has emerged as a new form of computing that protects data in use for workloads running in an untrusted environment, all the way from the cloud to edge devices. In developing Confidential Computing, Intel and other industry leaders formed the Confidential Computing Consortium, which brings together hardware vendors, cloud providers, and software developers to accelerate the adoption of Trusted Execution Environment (TEE) technologies and standards. The consortium concentrates on developing and promoting Confidential Computing as an emerging industry initiative focused on securing data while it's in use.

Before an application can process it, encrypted data must be unencrypted in memory. This can render that data vulnerable during processing on systems that are compromised, either externally or by malicious insiders. Confidential Computing helps to solve this issue by leveraging Intel's hardware-based Trusted Execution Environments, called Intel® Software Guard Extensions (Intel® SGX). These TEEs provide secure enclaves within the Intel® Xeon® Scalable processor. By using Intel® SGX, encrypted data can be processed in protected enclave memory with lowered risk of exposing it to the rest of the system. This helps reduce the vulnerability of sensitive data, while at the same time maintaining a high degree of transparency for users. Thanks to Intel® SGX, organizations can better enable these capabilities at scale. This is especially critical in multi-tenant cloud environments, where sensitive data must be kept isolated from even privileged areas of the system stack.

As computing moves to span multiple environments—from on-prem to public cloud to the edge—Confidential Computing powered by Intel® SGX will continue to play an increasingly important role in providing organizations with robust controls to help safeguard sensitive intellectual property and workload data, wherever that data resides—at rest, in flight, and during processing.

Gramine: streamlining strong security

A critical element in the ongoing development of Confidential Computing is Gramine (originally known as "Graphene"). Gramine is a library OS designed to run unmodifed Linux application binaries on other platforms. The current effort focuses on running Linux applications on Intel® SGX, but the project is intended to eventually support a variety of architectures and operating systems.

Gramine helps to greatly simplify and extend the use of Intel® SGX in Linux environments, and also simplifies running applications inside Intel® SGX enclaves. As a library OS that offers an alternative to the existing set of Confidential Computing SDKs, Gramine helps make it easier for IT professionals to deploy Confidential Computing solutions without code modification. It provides a "push button" method for more easily protecting applications and data, resulting in a faster, more secure, and more scalable end-to-end security solution with minimal effort.

Gramine is comparable to a lightweight operating system, and is designed to run single Linux applications in restricted environments. Gramine delivers a range of features that organizations benefit from, similar to running a complete operating system on a platform: multi-processing, multi-threading, networking, file system, etc.

In September 2021, the Confidential Computing Consortium announced that Gramine joined the Linux Foundation as an official Confidential Computing project:

"In untrusted cloud and edge deployments, there is a strong desire to shield the whole application from the rest of the infrastructure. Gramine supports this 'lift and shift' paradigm for bringing unmodified applications into Confidential Computing with Intel® SGX. Gramine can protect applications from a malicious system stack with minimal porting effort."

- https://confidentialcomputing.io/2021/10/08/gramine-1-0-release/

Cloud-native deployment

Gramine is cloud native and supports Docker container integration through the Gramine Shielded Containers (GSC) tool that automatically converts Docker images to Gramine images. Containers built with GSC can be deployed via Kubernetes for confidential containers and microservices. Gramine can also integrate with Azure Kubernetes services for deployment in Azure Confidential Computing Cloud, providing a very lightweight isolated environment compared to a complete guest operating system running inside a virtual machine.

Curated Applications with Gramine

Gramine delivers a growing set of curated applications, including modern Machine Learning (ML) frameworks like TensorFlow and PyTorch, database technologies like Redis, web servers, and programming language runtimes. These curated applications are tested and benchmarked, making it easier for developers to build security into their workloads from scratch. Utilizing Gramine curated applications enables organizations to more easily build full, secure, end-to-end solution architectures without concern about integrating frameworks.

Gramine Shielding Layer

As is typical in security-critical environments, the application running inside Gramine must be accompanied by a security manifest that defines the security posture of the application. Gramine uses that manifest to provide a shielding layer to help protect the application, based on the policies specified in the manifest. The manifest is the single most important file when enabling applications inside Gramine and Intel® SGX; ultimately, both the security and the proper functioning of the application depend on how the manifest file is written.

Attestation

When Gramine is used to create and run Intel® SGX enclaves, attestation is a critical part of that process. Attestation is a mechanism that verifies the integrity of the computing environment, enabling remote users to confirm that an application is running on real hardware in an up-to-date TEE. Gramine's built-in capabilities can help enable companies to verify compliance with security policies and SLAs. Gramine supports both Intel® Enhanced Privacy ID (Intel® EPID) and Software Guard Extensions Data Center Attestation Primitives (Intel® SGX DCAP)/Intel® Trust Domain Extension Data Center Attestation Primitives (Intel® TDX DCAP) attestations, currently the most common attestation protocols used in Confidential Computing.

Gramine use cases

Gramine is a flexible tool that can help address a wide variety of Confidential Computing use cases.

Securing the Cloud

The most common use of Gramine is to protect sensitive data while it's in use; this enables organizations to reap the benefits of cloud economics on all their data sets and workloads. When used in conjunction with the encryption of data in transit and at rest, Confidential Computing removes the biggest barriers to cloud adoption, freeing enterprises to move all their data sets to the cloud to move away from costly and inflexible infrastructure overheads. By enabling the easy movement of data, code, containers, and more to the public cloud, and protecting those assets while they are being processed, Gramine delivers a private cloud on the public cloud.

Privacy-Preserving Machine Learning

Machine learning is rapidly expanding today and is critical in an ever-growing number of real-world applications. For example, ML algorithms can analyze known past data to generate previously hidden results. Those results can then be used for a wide range of tasks, such as forecasting weather, classifying images, diagnosing illness, and many more.

But as Machine Learning pervades our daily lives, increasingly serious privacy concerns emerge. That's where Privacy-Preserving Machine Learning (PPML) comes into play. PPML protects the privacy of sensitive user data, as well as the ML-trained model (which itself is proprietary intellectual property). That helps enable the safe use of ML for critical tasks such as cryptography, hardware technologies, and differential privacy (where valuable information can be extracted from databases without violating the privacy of personal data contained in them).

- **Training with Private Data**
  Machine learning models are strongest when they are able to maximize their use of training data. Yet data privacy regulations often require companies to eliminate Personally Indentifiable Information (PII) data before training, which can yield weaker (i.e., lower confidence) inferences and elongate the training process, making it more expensive. Using Gramine, a machine learning application can access and analyze the PII data it needs while at the same time helping meet privacy regulations.
- **Inference with a Proprietary Model**
  A trained machine learning model is then used to perform inference on data. Frequently, this model itself is a highly-valuable proprietary asset of the company, and must be kept secret even during the inference process. Using Gramine for inference applications helps maintiain confidentiality of the model.
- **Inference with Private Data**
  Not only must the model be kept confidential during the inference process, but also the data. With Gramine, inference applications help guarantee confidentiality of the processed data.

Compliance

The processing of sensitive data sets in the public cloud requires compliance with security standards. Confidential Computing powered by Intel® SGX enables enterprises to use strong encryption and still adhere to compliance controls when they put their data to work. They can do this by decrypting data in enclaves, staying in control of keys, and confirming attestation to the integrity of the computing environment. As privacy technologies become ever more prevalent, Confidential Computing will continue to help enable privacy, driven by technology, in support of regulation. This helps data owners to significantly increase their compliance targets while reducing the overall cost of doing so.

Secure Collaboration

New compute paradigms are emerging to enable data sets and processing power to be shared across multiple parties. However, these data and computational models may be sensitive or regulated, as is often the case in the financial services, healthcare, government, and non-profit domains. Gramine provides the capability for companies to share data sets with each other for aggregation and cross analysis while remaining in control of their data. For example, private multi-party analytics can be applied when multiple parties have private data that needs to be combined and analyzed without exposing the underlying data to any of the other parties. This technology can be applied in multiple ways, such as the prevention of fraud in financial services, the detection and development of cures for diseases in the healthcare industry, and many more.

Gramine can help organizations ensure that data on remote systems is protected against tampering and compromise (including against insider threats within the partnering organizations) and can also help validate the integrity of the code processing that data. In addition, data can be combined and analyzed within Intel® SGX, and the results can be sent in an encrypted format back to each party. That means the data remains protected throughout the entire process: in transit, during computation, and while stored. These capabilities will help drive the advancement of a global data-sharing playing field where organizations can unlock previously unleveraged data sets for collaborative analysis and exchange with other organizations—and do so with reduced risk of security, privacy, and regulatory impacts.

Secure Smart Contracts for Blockchain

By combining the power of Confidential Computing and Blockchain technologies, users can leverage Intel® SGX to provide attestation and verification services that help optimize scalability, privacy, and security.

Assuring data consistency among Blockchain users usually means each party must independently validate historical data (upon which the validity of any current data depends). This requires all parties to have visibility to those historical data sets, and that raises potential scalability or privacy concerns. To deal with this, users can execute smart contracts inside an Intel® SGX enclave instead of independently accessing and validating historical data and associated smart contracts for themselves. Once such a transaction is finalized, Gramine can provide attestation services to help confirm its reliability, which helps free subsequent participants from the need to perform the verification again for themselves. In addition, Gramine-based attestation services can also help address some of the computational and communication inefficiencies that arise from consensus protocols.

Gramine story and community

Gramine (originally Graphene) started as a research project at Stony Brook University in 2011 and was first released in 2014. Intel® SGX support was added through co-development with Intel Labs, and the first Intel® SGX version was released in 2015 and later stabilized and published in 2017.

The Gramine open source community was formed in early 2019, when Invisible Things Lab and Golem joined hands with Intel and academic partners to create a community project. Since then, the development community has grown into a diverse group of individuals, developers from both large and small companies, and university researchers.

In the fall of 2021, the Graphene project joined the Confidential Computing Consortium under the new name Gramine.

Contributors to the project are growing in number, as are their products. Services and applications built on top of Gramine are delivered for a variety of industries, from data-sensitive sectors such as health care and financial services to highly scalable areas with large data sets, such as the industrial and telecom sectors.

Intel partners such as Cosmian are utilizing Gramine to leverage external Machine Learning libraries. Taking advantage of Intel® SGX and Gramine allowed Cosmian to develop their scalable Cosmian Secure Computation API-based solution, which allows encrypted Python code to run over encrypted input data without revealing anything to the cloud.

According to Bruno Grieder, CTO of Cosmian:

"Gramine is very powerful to enable collaborative, confidential computations. This is particularly critical in the healthcare sector, where biotechs, pharma labs, and hospitals need to analyze sensitive medical data without directly processing data in clear text."

In addition, the Swiss university Fernfachhochschule Schweiz (FFHS) has developed a fraud detection mechanism for distance learning exams that incorporates video inputs from exams using Intel® Distribution of OpenVINO™ toolkit. Due to the privacy aspects of the video capture, they were required to ensure that the video was always encrypted. To remain compliant, they now run their OpenVINO instances inside Gramine.

Collaboration with Confidential Computing Open Source Community

Intel is also committed to working with confidential computing open source partners to ensure the growth, enhanced usability, and scalability of Gramine. Partners such as Edgeless Systems have shown how scaling and orchestrating of Gramine in a cloud native environment can be done. Edgeless Systems also offers a production-ready backend for attestation verification and key management for Gramine workloads.

Gramine Features	
SGX attestation	Transparent attestation support proves to remote verifier that expected code execution took place in genuine Intel® SGX enclaves.
Protected files	Security-critical files can be automatically encrypted and decrypted inside the enclave, protecting from malicious attacks.
Curated applications	Pre-integrated and benchmarked applications available for Gramine via GitHub. Examples include TensorFlow, Redis, PyTorch, and many others.
Application-oriented security	Gramine helps ensure secure execution of the application even in untrusted environments.
Asynchronous system calls	Exit-less system calls help improve performance of critical applications.
Multi-process support	Support for multi-process applications through full fork support, establishing a protected channel for local attestation. Gramine is one of few frameworks offering this.
Confidential containers	Easy deployment with confidential containers. Gramine Shielded Containers (GSC) tool automatically converts Docker images to Gramine-shielded confidential containers.

Learn more

Gramine Github repository
https://github.com/gramineproject/gramine
Gramine Project
http://www.gramineproject.io
Gramine Documentation
https://Gramine.readthedocs.io/en/latest
Gramine Support Group
https://gitter.im/gramineproject/community
Confidential Computing Consortium
confidentialcomputing.io/2021/10/08/gramine-1-0-release/
Intel Newsroom
https://www.intel.com/content/www/us/en/newsroom/news/computing-consortium-announces-gramine-1-0.html?wapkw=gramine#gs.mgthpp
Gamine for Machine Learning: A Tutorial
https://gramine.readthedocs.io/en/latest/tutorials/pytorch/index.html

ACG6228ISS