SOLUTION BRIEF

Enterprise and Government Security

An Introduction to Confidential Computing

Secure compute in the cloud era with Intel® Software Guard Extensions

Cybersecurity is the number one concern of CIOs.[1] But as enterprise IT architectures become more distributed and complex, the task of maximizing security becomes increasingly challenging. A range of technologies exist to help protect data in flight across the network, or at rest in storage, but until recently, options have been limited for protecting data in use while it is being actively processed in memory. That data is nearly always unencrypted and potentially vulnerable to privileged attacks from malware and/or malicious insiders.

This paper is about technological solutions to help protect that data, collectively called Confidential Computing. It covers some of the use cases which this new paradigm of security empowers, and the consortium of industry leaders who have come together to realize this vision.

Public cloud increases vulnerability

So, why now? This potential vulnerability has always existed without being seen as a major issue. One reason is the growth of cloud computing, and especially the public cloud. Enterprises are running more and more workloads on infrastructure that they do not themselves control, raising both practical and compliance concerns about the security and alterability of data and applications. Threats could include malicious system administrators or other insiders at the Cloud Service Provider (CSP), hackers exploiting bugs in the CSP's cloud fabric, or other third parties accessing data without customer consent.

Additionally, internet of things (IoT) data is increasingly processed, analyzed and sorted at the edge to address either latency or network data volume limitations. This data also needs to be better protected in an environment which is typically less easily managed and controlled than a data center environment.

Trusted execution environments are the basis of confidential computing

Confidential computing is a stack of hardware and software which together work to address core security concerns of the cloud age – from data security in the public cloud to edge security which enables federated learning and accelerated blockchain.

[1] https://www.techrepublic.com/article/digital-transformation-top-5-concerns-for-cios/

At the base of that stack is the trusted execution environment (TEE) - also called an enclave - where data and code are isolated and shielded from other software, including the operating system and cloud service stack. The hardware protects a portion of the processor and memory, on which only authorized code is permitted to run and to access data, so code and data are protected against viewing and modification from outside of the TEE, even with privileged root access.

This is done by a combination of physically encrypting a portion of memory, and changing the memory access control so that previously privileged software (OS, hypervisor etc.) can no longer access or 'see' the data or application code within that enclave. Developers can use libraries and extensions such as Intel® Software Guard Extensions (Intel® SGX) to create programs which use these enclaves. In this way the TEE provides four key benefits:

- Strong protection from both on-chip and off-chip unauthorized accesses
- No reliance on privileged software managed by administrators
- Verification that the right code is running on the right hardware
- A common development and deployment model.

Figure 1 shows how Intel SGX enclaves work in practice. An application is built with trusted and untrusted parts.

That application runs and creates a special protected portion (the trusted enclave) with restricted entry/exit location defined by the developer. When the trusted function is called, only the code running inside the enclave sees data in the clear (decrypted). Enclave code and data inside the CPU perimeter run in the clear, and enclave data written to memory is encrypted and integrity checked. All external access to enclave data is denied. The function returns, and the enclave data remains in the trusted memory space. This provides assurances that the data within the enclave remains confidential, unmodified, and more secure.

Use cases for confidential computing

More secure public cloud infrastructure

Enterprises remain unsure whether to migrate some sensitive workloads into the public cloud because of security and privacy concerns. Confidential computing substantially mitigates the risk of snooping on, or alteration of, sensitive data in the public cloud environment. This gives enterprises the flexibility to confidently place more and different workloads into the public cloud. Data is always in the control of the customer even when the infrastructure on which it runs is not, and is opaque to the cloud platform or its administrators.

Several CSPs, including Microsoft Azure, IBM Cloud, Alibaba Cloud, and more already have confidential computing environments based on Intel SGX.

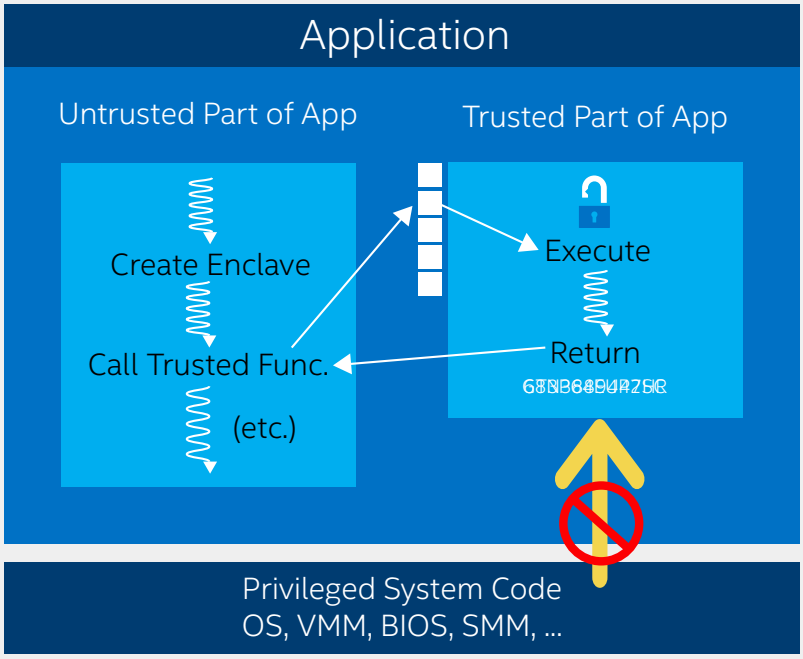

Figure 1. Running trusted enclaves with Intel® SGX

Enterprise Blockchain with performance and privacy

One of the established principles of blockchain is that the data and validation logic for every transaction happens 'on chain', at every node of the blockchain network. This is what makes the distributed ledgers difficult to hack, corrupt, or alter. Nevertheless, this model does come with some issues, primarily around scalability and performance. Essentially every node on the chain is a verifier. This is very computationally expensive. Furthermore, there may be situations where you want to keep the contents of the blockchain contract private and only exposed to the direct parties to that contract (not the verifiers).

One option which confidential computing makes possible to mitigate these issues is to take some of the operations 'off chain'. This risks reducing the resilience and security of the blockchain, but the use of confidential computing overcomes these concerns, and opens the way to significantly reducing scalability and performance challenges without unacceptably compromising the blockchain.

Using Intel SGX to encrypt the ledger helps ensure that only authorized parties can see the transactions on the network. In this way, blockchain data is in encrypted form until it is needed for a transaction, then decrypted in an enclave, where only permitted participants are allowed to view it.

The use of TEEs in a design for a trusted distributed blockchain has been codified in the Confidential Consortium Blockchain Framework. The framework simplifies consensus and transaction processing for high throughput and fine-grained confidentiality (see Figure 2).

There are also other projects building on trusted computing for blockchain, for example, the Hyperledger Avalon project. This hyperledger project brings together sponsorship from Intel, iExec Blockchain Tech, Alibaba Cloud, Baidu, BGI, Chainlink, Consensys, Enterprise Ethereum Aliance (EEA), Espeo, IBM, Kaleido, Microsoft, Banco Santander, Wipro, Oracle, and Monax. The project builds on the EEA's Off-Chain Trusted Compute Specification, which standardizes distribution and reconcillation of workloads.

Secured federated learning

Federated learning is when multiple data sources (often located within different companies or organizations) are used to train a single machine learning model, which is applied to each data source independently. The data host computes an update to the current model based on its local data, and communicates this update to a central server, where these updates are aggregated. This avoids the bandwidth, risk and security problems of transmitting large amounts of confidential or valuable data to a central source. However, it does have a number of potential security drawbacks. The first is that neither the data nor the model are inherently protected. The model can be corrupted or 'poisoned' and the data potentially viewed or altered without authorization.

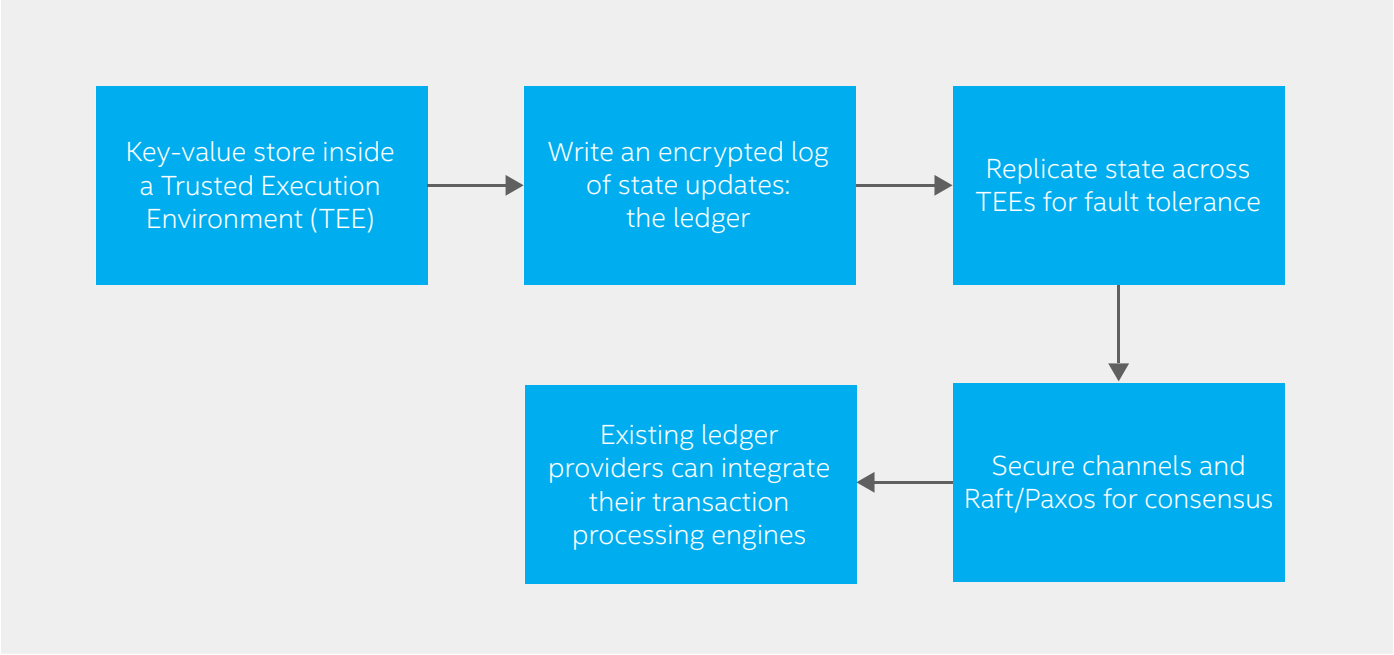

Figure 2. Confidential Consortium Blockchain Framework (CCBF) Design

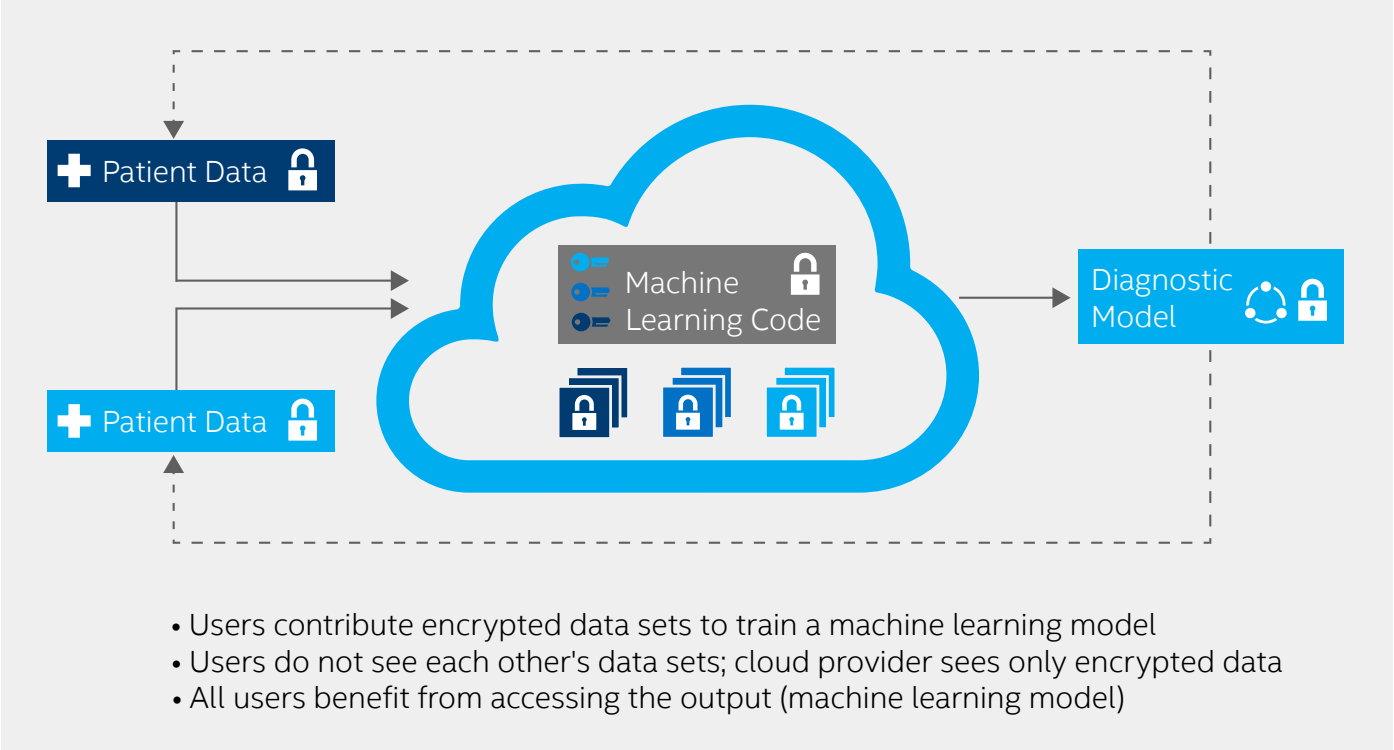

Figure 3. Secured federated learning, healthcare use case

This vulnerability is very challenging for a number of potential use cases, such as training a model on patient data across multiple hospitals to create better diagnostic models; rival banks working together to create better anti-money laundering models; or retailers and commercial partners working together to make more targeted offers (for example an airline and a credit card company). In all of these cases the potential benefits are huge, but there needs to be a way of protecting both the sensitive data and the model.

Confidential computing allows exactly that. Both the algorithm and the data sets are encrypted and better protected. With the data analysis running in protected enclaves, the data and the model are designed to avoid exposure to potential interference. The idea is that only approved models and procedures are used and that the collective rules are respected. Enclaves are designed so that none can see the other's data, nor can the central owner of the model. Likewise, the intellectual property of the model itself also receives protections. Figure 3 shows how this works in the medical example above.

Secure networking

Operators need to develop, provision and support 5G and next-generation networks efficiently and cost-effectively. This is driving a move away from bespoke single-function equipment towards running network functions and business support systems (BSS) in a cloud environment.

There are currently trials ongoing by multiple operators to see how confidential computing can secure their BSS, including billing systems and keys to protect virtual network functions (VNFs), substantially lower auditing costs and total cost of ownership (TCO).

Furthermore, as network operators are pushing more compute towards the edge they encounter essentially the same security issues as the cloud providers described above. Confidential computing offers real scope to mitigate these concerns and offer reassurance to data owners.

The Confidential Computing Consortium and Intel's contribution

Intel and other industry leaders have come together to form a Confidential Computing Consortium under the Linux Foundation. We're proud to be a founding member of this new industry group dedicated to making confidential computing practices, such as the protection of data in-use, easier to adopt in today's multi-cloud world.

Mark Russinovich, CTO of Microsoft Azure sums up the value of the Confidential Computing Consortium perfectly: "These technologies offer the promise to protect data and enable collaboration to make the world more secure and unlock multi-party innovations."

"Federated learning distributes the machine learning process over to the edge. It is a new framework for artificial intelligence (AI) model development that is distributed over millions of mobile devices, provides highly personalized models and does not compromise the user privacy. The model development, training, and evaluation [are executed] with no direct access to or labeling of raw user data".

—Dr. Santanu Bhattacharya[2]

The Confidential Computing Consortium is initially focused on common programing models and enclave portability, but the Consortium doesn't prescribe the hardware mechanism necessary for creating and protecting the enclave. That's where Intel SGX comes in.

Intel SGX is a hardware-based technology that helps protect data in use by establishing protected enclaves in memory, so only authorized application code can access sensitive data. Unlike full memory encryption technologies that leave the data within the attack surface of the operating system and cloud stack, Intel SGX allows a specific application to create its own protected enclave with a direct interface to the hardware, limiting access and minimizing the overall performance impact for both the application and any other virtual machines (VMs) or tenants on the server.

Intel SGX provides hardware-based encryption for data in use protection at the application level with the smallest attack surface. Intel SGX is available today on Intel® Xeon® E-2100 processors, and by using the Intel® SGX Card, a PCI-Express add-in card, that enables Intel SGX in multi-socket Intel® Xeon® Scalable processor servers. Intel SGX will continue to be rolled out across upcoming generations of mainstream Intel Xeon platforms.

Intel SGX is used in confidential computing services from Microsoft Azure, IBM Cloud Data Guard, Baidu, Alibaba Cloud and Equinix.

Further reading

- **The Confidential Computing Consortium:** https://confidentialcomputing.io/
- **Intel Security:** https://www.intel.com/content/www/us/en/security/hardware/hardware-security-overview.html
- **Intel® Software Guard Extensions (Intel® SGX):** https://software.intel.com/en-us/sgx https://software.intel.com/sites/default/files/managed/c3/8b/intel-sgx-product-brief-2019.pdf

Solution Provided By:

[2] https://towardsdatascience.com/the-new-dawn-of-ai-federated-learning-8ccd9ed7fc3a

1219/LD/CAT/PDF 341265-001EN