White Paper

The Enclave Device Blueprint for Confidential Computing at the Edge:

A community approach to bring confidential computing at the edge and IoT into mainstream for privacy and safety

A White Paper by Arm, Microsoft and Scalys

Authors:
Eustace Asanghanwa, Principal Program Manager, Microsoft
Femi Idowu, Cloud Solutions Architect, Microsoft
Evert Pap, Architect and Technical Lead Engineering, Scalys (Subsidiary of Sintecs BV)
Dmitry Lavnikevich, Software Architect, Scalys (Subsidiary of Sintecs BV)
Paul Howard, Principal System Solutions Architect, Arm
Marc Meunier, Sr Manager Software Ecosystem Development, Arm and Governing Board Member on Confidential Compute Consortium

To help protect data in the Internet of Things (IoT), developers are using a relatively new technique from the cloud, called confidential computing, to isolate sensitive operations in a Trusted Execution Environments during processing. IoT data remains protected when in use, so it's easier to meet strict regulatory requirements for data privacy and introduce new use cases involving sensitive data. This paper summarizes confidential computing and its potential for edge and highlights how Arm and Microsoft are working with other industry leaders to make it easier to enable secure and isolated environments for confidential computing.

Securing the Cloud was the first step

Cloud computing is an essential infrastructure for most enterprises and developers today. Gartner expects global end-user spending on Public Cloud Services to exceed $480 Billion in 2022. Today, majority of the data produced by enterprises and endpoint devices is processed in the cloud, and with more and more information being transmitted to, stored in, and processed in hybrid, public, and multi-cloud environments. With so much data at risk, the issue of cloud-based data security becomes increasingly important. Regulatory bodies have also taken notice of information leaks and are now more active in defining requirements around data privacy.

When considering how best to protect data in the cloud, it helps to begin by thinking about data states, since the state of cloud data – that is, what data is doing at any given moment – can influence the security requirements.

In general, there are three data states to consider: at rest, in transit, and in use. When data is at rest, it is not being accessed and is stored somewhere, such as a file server or a database. When data is in transit, it is using a network connection, such as Wi-Fi, cellular, or Ethernet, to travel from one point to another. And when data is in use, it is actively being processed by application code, as part of a workload during runtime.

Firewalls and encryption have been used for decades to protect data while it's at rest and in transit. Firewalls, for example, are a first line of defense, preventing unauthorized access to the network and isolating one network from another. Encryption, which scrambles data and renders it unreadable to anyone without the correct decryption key, is another powerful security mechanism, since it ensures that data is safe even if attackers get past the firewall. Any encrypted data obtained by an unauthorized person or entity is unreadable and essentially worthless.

Protecting data in use, however, is something that has been harder to do. That's because, with standard technology, data must first be decrypted before it can be processed. The moments before, during, and after processing, when data is unencrypted, are vulnerable times that offer opportunities to perform a memory dump, compromise the root user, or carry out other malicious acts.

> Confidential computing fills a security gap by protecting data while it's in use

A recently developed technique for cloud operations, called "confidential computing," helps fill this security gap by protecting cloud data while it's in use.

Protected Processing in the Cloud

Confidential computing secures data in use by isolating it in a hardware-based Trusted Execution Environment (TEE). This protects the code and data within the environment and prevents unauthorized access.

While data is being processed, it's invisible and unknowable to the operating system, the hypervisor of a virtual machine, and other compute-stack resources. It also can't be seen by the cloud provider or their employees.

Confidential computing makes it safer to run sensitive applications on public clouds

Extensive Cloud Benefits

By protecting data in use, confidential computing makes cloud-based operation safer, more attractive, and more flexible. Confidential computing gives greater assurance that cloud operations are protected and private, and that makes it safer to move more applications – including those that involve sensitive data, such as Personally Identifiable Information (PII), financial details, and medical records – onto hybrid and public clouds. Sharing multi-tenant environments become less risky, too, because there's less chance that workloads from one tenant will interfere with workloads of another, or that a security breach will allow a tenant to gain access to another tenant's data or resources.

Confidential computing isolates code, along with data, so it's easier to protect Intellectual Property (IP), such as proprietary business logic, analytics functions, algorithms that use machine language or artificial intelligence, and even entire applications. At the same time, having the ability to retain control of private information makes it easier to collaborate with third parties, without risking proprietary data or algorithms. Each party shares only what it needs to for the project, and is able to gain insights from the collaboration, while sensitive information remains beyond the reach of those who don't need to see it.

Another benefit to cloud customers is that confidential computing makes it possible to avoid vendor lock-ins, which prevent users from switching services. Businesses can use a mix of providers, choosing the best one for each technology or service, and don't need to worry if the competition uses the same provider.

A Long-term Trend

Confidential computing is still a comparatively new feature of cloud services but, due to its many business benefits, has caught on quickly. It promises to be a lasting trend. Major cloud service providers like Microsoft Azure now offer it as an option, and rapid expansion is expected across all geographic regions. The analysts at Market Digits Forecast predict the global market for confidential computing will have a Compound Annual Growth Rate (CAGR) of 52.21 percent between 2022 and 2030, for a Total Available Market (TAM) of USD 37 billion.

Another indication that confidential computing is an important technique with a promising future is the fact that it now has its own industry group. The Confidential Computing Consortium (CCC) brings together hardware vendors, cloud providers, and software developers to accelerate the adoption of TEE technologies and standards. The CCC fosters cross-industry collaboration in an open framework, and includes a project community at the Linux Foundation. Arm and Microsoft are premier members of the CCC and actively engaged in the CCC's efforts to define and accelerate the adoption of confidential computing.

Edge Migration

The proliferation of IoT devices has introduced new challenges around data processing. In many cases, there is so much data being generated at the Edge that it is no longer feasible to send all this data up to the cloud for processing. In many cases today, data needs to be processed at the Edge and only a subset of the data makes it to the cloud. Intelligent edge devices often process highly sensitive workloads to generate results that are also highly sensitive, making it necessary to use strong protection to preserve integrity, privacy, and confidentiality. IoT devices may not always be in trusted custody, and have become frequent targets for tampering, forgery, and other types of hacking. To add to the challenge, devices at the edge may be connected to the cloud, but might also need to operate independently from the cloud connection.

The increasing use of intelligent edge devices, equipped with more processing power than the typical sensor or actuator, makes it all the more important to protect IoT applications.

Public clouds are also a prominent part of the IoT picture, since today's IoT companies tend to use a cloud-only approach that employs someone else's IT infrastructure. The IoT devices in these deployments are, by and large, using cloud-based solutions to handle security-sensitive data and software, so it's important for IoT companies to understand how confidential computing can help secure backend systems, comply with data-residency rules, and protect privacy.

But IoT deployments can also benefit from confidential computing in another way. Many of today's edge and IoT devices can be thought of as cloud computers in miniature, with the same need to protect data at rest, in transit, and in use. This is especially true for edge devices that process and analyze data before sending it to a gateway or cloud, and for IoT devices equipped with storage for housing collected data. By migrating confidential computing from the cloud to the edge and into IoT devices, it's possible to create end-to-end protections that safeguard data at every point, from silicon to the cloud.

> Edge devices are popular targets for hacking

Benefits of Confidential Computing at the Edge

- Secure execution of code – run only sanctioned, intact code and control access to resources
- Trusted I/O for command and control – protect the command centers of critical infrastructures
- Secure reporting – compose secure reports, for use in billing and other areas
- Secure metering – deploy consumption-based monetization or warranty management
- Secure activity logging – enable high integrity recording of activity for non-repudiation
- Protected data confidentiality and integrity – defend against hackers and malware that exploit bugs in the operating system

Protecting Edge Data and Outcomes

When deployed in an edge device, a Trusted Execution Environment (TEE) provides the same kind of protected environment for processing. Sensitive data and operations used in the IoT device can't be observed by other pieces of the system, making it extremely difficult to modify data, subvert code, or influence outcomes.

Creating a secured TEE in device hardware provides a higher level of security than using software alone to protect data. At present, the edge device places a high level of trust in supervisor software, in the form of kernels and hypervisors that manage applications and virtual machines. The drawback, from a security standpoint, is that supervisors can access the data and code used by applications, and hackers can manipulate supervisors to leak confidential data or algorithms held in applications. With confidential computing implemented on the IoT device with hardware-based TEEs, the supervisor's access status changes from "will not" to "cannot," thereby minimizing a critical vulnerability and essentially eliminating one of the ways that hackers exploit deployments for exfiltration and malicious tampering.

What's more, adding confidential computing to an IoT device makes it possible to protect offline availability, when device isn't connected to the cloud but still need to process data locally, such as when devices are deployed in hard-to-reach areas, such as an agricultural field, a utility grid, a pipeline, or on an offshore ship. Offline availability also covers situations where privacy is so important that data should not leave the system, such as in medical equipment and high-security facilities.

Deploying edge devices that use confidential computing – by use of a hardware-based TEE– opens up possibilities for new use cases, especially those that involve sensitive data or generate insights that need to be protected. For example, medical researchers can use TEE devices to collect health data without risking personal identities, and manufacturing facilities can use machine-learning algorithms to control factory equipment while safeguarding proprietary techniques. TEE devices also make it easier to work with proprietary algorithms, used to derive unique insights, while minimizing the risk of revealing proprietary methods and losing a competitive advantage. TEE devices can also make it easier to support the zero-trust model of computing, where software runs on someone else's hardware, and the model of software offloading, where devices locate and use nearby compute resources.

Developing with Trusted Execution Environments at the Edge

In the cloud, confidential computing is a feature that cloud providers build and operate on behalf of their customers. But developers of edge devices, looking to add confidential computing to their design, need to build their own solutions, and that can be challenging and expensive. Bringing confidential computing to the edge involves a new approach, both in how developers think about edge security and how they create edge devices.

That being said, the idea of using hardware based security creating a protected Trusted Execution Environment in hardware, to safeguard sensitive data and operations at the edge, isn't actually new. Banking, government, and other niche industries have used dedicated hardware chips like HSMs and TPMs to protect financial transactions, personal identities, security clearances, and other sensitive information for many years. For a long time, though, the development of applications that use HSMs was something that only a few experts, with extensive experience in tamper proofing and specific HSM hardware offerings, were equipped to do. Delivering a secured application with a HSM at its center meant having direct access to the HSM hardware, and involved a time-consuming, rigorous process to verify protection. These devices are also quite limited in what they can process and are used more to protect secrets then to protect operations.

> For many years, only experts in hardware security had the skills needed to work with HSMs

In recent years, TEE functionality has expanded to support flexible use cases and there's been a concerted effort to collaboratively expand the ecosystem for secure applications in the IoT. The Confidential Computing Consortium is just one example of how the developer community is coming together to make confidential computing more accessible.

The result is that developers now have access to a growing number of building blocks that support the emulation of TEEs and the development of confidential applications. That means that developing and debugging trusted applications is no longer a specialty, reserved only for TEE hardware engineers, but something any developer can take on.

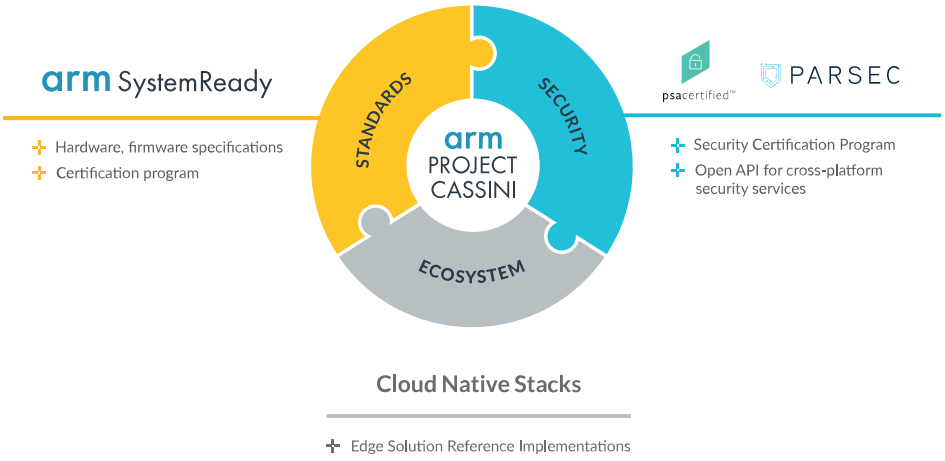
A Modular Approach

Through a collaboration with the IoT team at the cloud provider Microsoft Azure, the developers at Scalys, a frontrunner in the development of edge security, and Arm, we have created a modular approach for developing confidential compute capable devices. The approach uses Trust Zone paired with Arm Cassini initiatives to provide a base of security in the Azure IoT Edge development machine, the Open Enclave Software Development Kit (SDK), and the Azure-certified TrustBox Edge from Scalys.

> We want to make it easier for everyone in the IoT to take advantage of confidential computing

Arm Project Cassini

Arm Project Cassini is an open collaborative framework that pulls together standards and best practice to ensure a secure and cloud native software experience that scales across Arm devices.



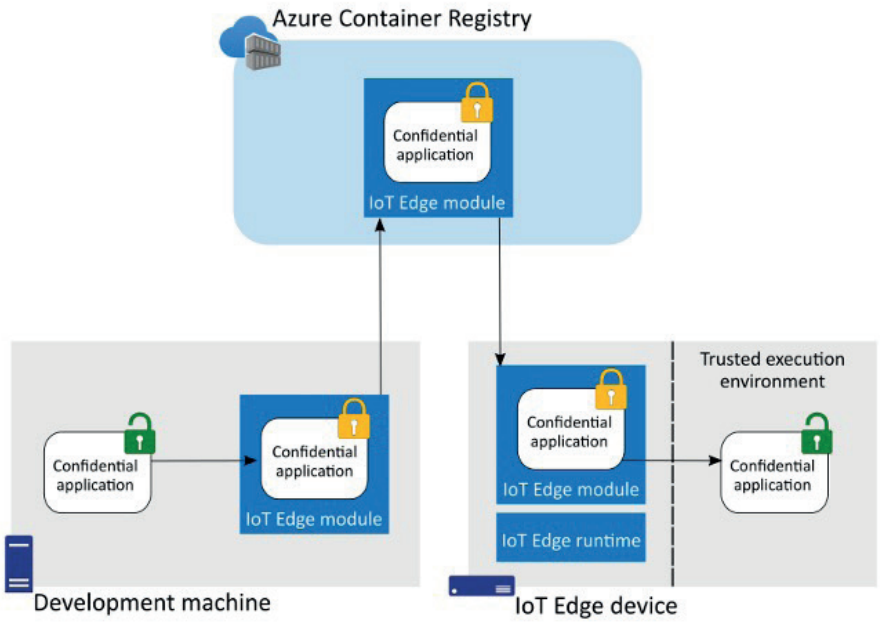
Azure IoT Edge

Azure IoT Edge is a fully managed service from Microsoft that delivers cloud intelligence locally by deploying Azure AI, Azure services, and custom logic directly onto edge devices, including gateways. This lets organizations bring AI and cloud analytics to the edge, especially when connectivity is poor or not available, or when high latency or high costs prevent connection to the cloud.

In a single development environment, developers can create confidential applications for deployment in cloud Trusted Execution Environments and in TEE-enabled IoT edge devices. The Azure IoT Edge enables safe and secured execution inside of Trusted Execution Environments. The workloads that execute in Trusted Execution Environments, referred to as confidential applications, are encrypted from the time they leave the developer build machine to when they land inside the device's TEE, where they are decrypted for safe execution.

As shown in the diagram, once a confidential application is created in the development machine, it is packaged as an IoT Edge Module. The application is encrypted before being pushed to the container registry, and remains encrypted, throughout the IoT deployment process, until the module is launched on the IoT Edge device.

Not until the confidential application is within the confines of the device TEE is it decrypted and available for execution. An additional layer of asymmetric encryption is used to transfer data confidentially into the TEE of the target device without it ever being exposed outside of the TEE. The private key is stored within the TEE and never revealed.



Confidential Data and Code Remains Encrypted Until Executed in the TEE
(Source: Microsoft, 2020)

Azure Samples

Azure Samples is an official Microsoft repository that provides a centralized experience for code-sample discovery, with the following goals in mind:

- A centralized location for both developers and customers to discover samples for any Microsoft tool, product or service.
- A flexible means to access source-control on GitHub, keeping pace with ongoing community-contributions and empowering the developer experience.
- An accessible web frontend that promotes code discovery and distribution, as well as collaboration between partners.

Azure Samples is organized as a repository with three basic building-blocks:

- The backend service: Leverages index-driven searches across code-sample repository hosted on Github to ensure that customers have a direct access to the latest version of code-samples.
- The Samples GitHub Code-Repository: Individual repositories that contain code-samples hosted on GitHub.
- Sample Browser: The customer-facing front-end experience situated on the Azure Samples project home page. This is the central page where customers can find relevant code samples through tailored searches.

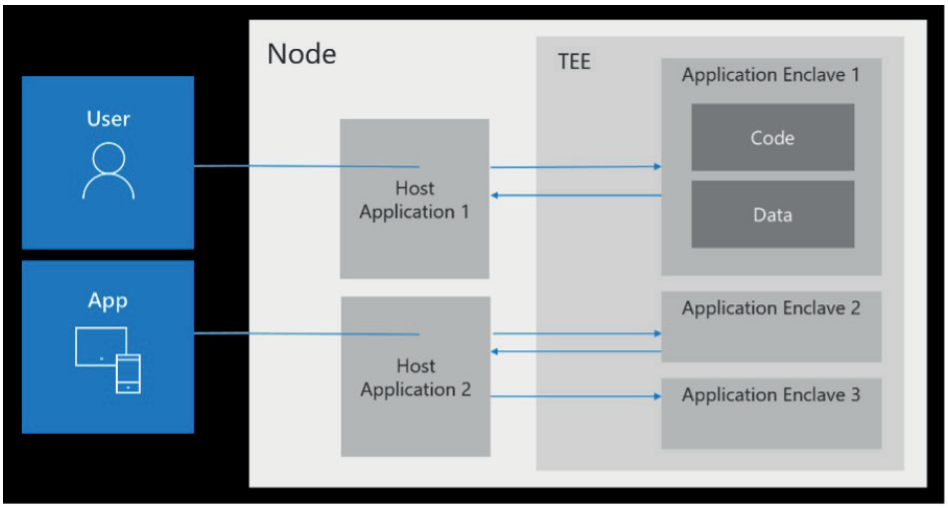
Code-base

It is important to note that the Code-base located on Azure-Samples, is code that is meant to provide examples, programs, and libraries, with the purpose of educating developers on how to build basic functionality, that enables them to eventually create more complex applications.

Open Enclave SDK

The Open Enclave SDK is, as its name implies, an open-source library for developing applications that make use of hardware-based TEEs, or Trusted Execution Environments. Launched by Microsoft in 2018 and now governed by the Confidential Computing Consortium, the Open Enclave SDK is a hardware-agnostic tool for building confidential applications for multiple platforms and environments in C and C++. It provides a single abstraction layer over any hardware-based TEE.

The resulting application partitions itself into two components: an untrusted component, called the host, and a trusted component, called the Enclaves. The host component runs unmodified on the untrusted operating system, while the trusted component runs within the TEE.



A single TEE can support multiple Confidential applications
Source: OpenEnclave.io, 2021

The SDK generalizes the development of confidential applications across TEEs from different hardware vendors, and is created with multiple software platforms, including Windows and Linux, in mind. In its present implementation, it provides support for Intel SGX as well as preview support for OP-TEE on Arm TrustZone.

As an open-source project, Open Enclave SDK is agnostic to specific vendors, service providers, and choice of operating systems. The project encourages community engagement and provides links, from its GitHub page, to information about how to contribute.

Scalys TrustBox Edge

The Scalys TrustBox Edge is an industrial-grade, tamper-resistant, secure router and IoT gateway that is certified for Azure IoT Edge and features a collaborative integration of Arm TrustZone security technology, NXP Layerscape processing, and the Open Enclave SDK.

Winner of the 2019 CES Best of Innovations award in the category of Cybersecurity and Personal Privacy, the TrustBox Edge is optimized for confidential computing usingTEE and lets developers focus immediately on the creation of confidential workloads.

In its latest iteration, the TrustBox Edge design is based on the NXP Layerscape LS1028A dual-core processor, a low-power communications processor with a 64-bit Arm Cortex-A72 core. The LS1028A features an Arm TrustZone, a hybrid hardware and software mechanism for protecting sensitive assets that supports TEE implementations. The design comes pre-loaded with the Open Enclave SDK and is Azure IoT Edge Certified as well as in the process to meet Arm SystemReady certification.

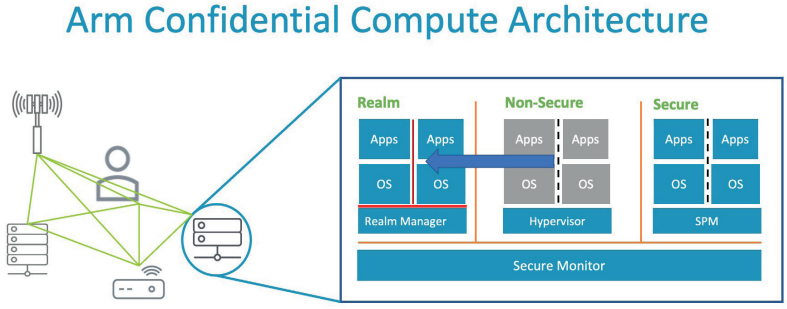
Using the TrustBox Edge, development teams can save time and effort by evaluating on the same hardware as goes to production. The design already meets industry standards for security and compliance, so development can proceed directly from evaluation to pilot and production.

Migrating to the Silicon Level

Arm is working to expand the confidential compute capabilities by making it a more integral part of the compute infrastructure with Arm compute architecture. The Armv9 architecture, which is Arm's first major chip architecture upgrade in a decade, includes a Confidential Compute Architecture (CCA). The CCA protects in-use portions of code and data from access or modification, even from authorized software, by putting computation into hardware-based security.

The CCA uses a concept called Realms, which are similar to containers and serve a purpose similar to a TEE. Realms work in a region separate from secure and non-secure environments. They protect sensitive data and code, whether the data is at rest, in transit, or in use. The Arm CCA builds on the strong security foundations of Arm TrustZone and will extend the broad adoption of confidential computing to every industry sector where microprocessors are in use.

With Armv9, confidential computing becomes an integral part of the compute architecture



Confidential Computing Concepts are Central to the new Armv9 Architecture (Source: Arm 2021)

Enclave Device Blueprint

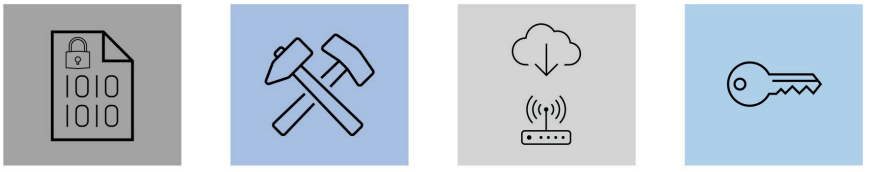
The Enclave Device Blueprint is a collection of projects, technologies, services, and motions to help realize a scalable end-to-end secure IoT edge deployment with confidential computing.

The Blueprint aims to solve specific challenges:

- Simplify enclave device engineering: abstract and simplify beyond Open Enclave SDK
- Enable enclave devices to scale across different technologies and OS by enabling OEMs to engage at a higher level of abstraction.
- Provide tooling resources to enable building end-to-end solutions. eg. resources needed in a developer build machine to create the confidential application workloads and workflows
- Enable maintenance and support scale by contributing and building from a common OSS project. Solution builders are continually asking for 10 years or greater device maintenance support.

Essential Blueprint components:

There are new components required to create the secured end to end framework. These include:



Packaging

The blueprint introduces a new portable file format specification: the **Confdential Package (.cpk)** file, which embeds encrypted applications for secure storage and transport.

Tooling

The new **cpk-tool** shell command is used to assemble confidential packages in your CI pipeline, and to install them on a target device. Sign within the tool or using offline signatures. Certs can be embedded or referenced.

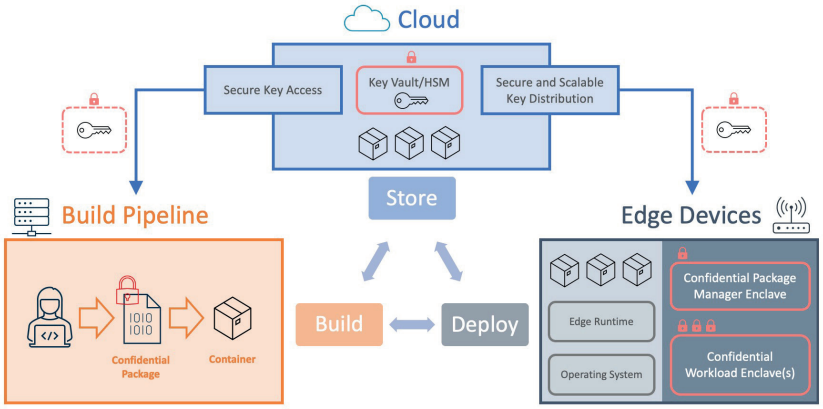
Installation

The **Confidential Package Manager** runs in a secure enclave on the target device. The **cpk-tool** interacts with this component to enable the secure delivery of workloads into new enclaves on the device.

Key Management

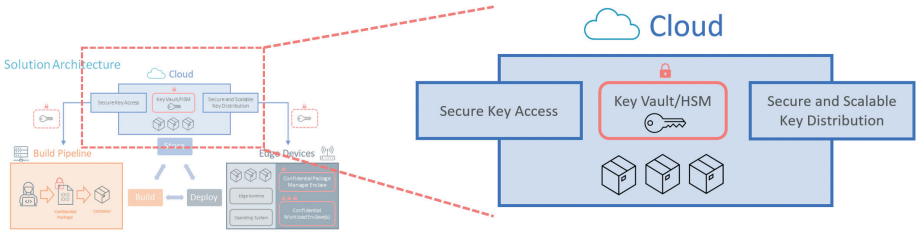
The blueprint includes **flexible and scalable contracts** for end-to-end key management across the build pipeline, the cloud, and the target device. Supports BYOK mechanisms. Distribution via proven scalable channels.

Solution Architecture



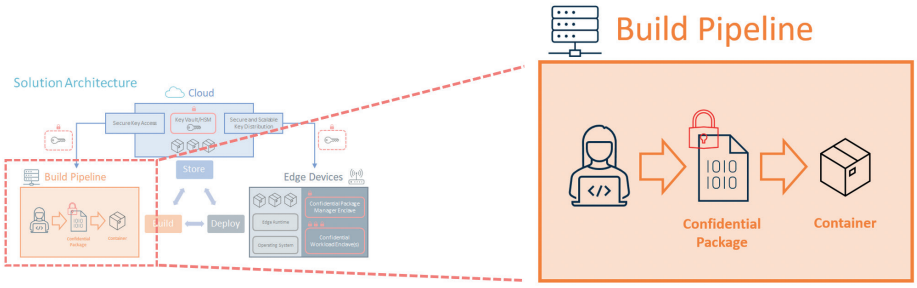
The overall architecture covers the end-to-end solution that spans across the build pipeline, across the cloud infrastructure and the end devices with their confidential enclaves. The goal of the solution is to build a confidential application in the build pipeline and ensure that it is never exposed outside the secure environment.

Key management framework



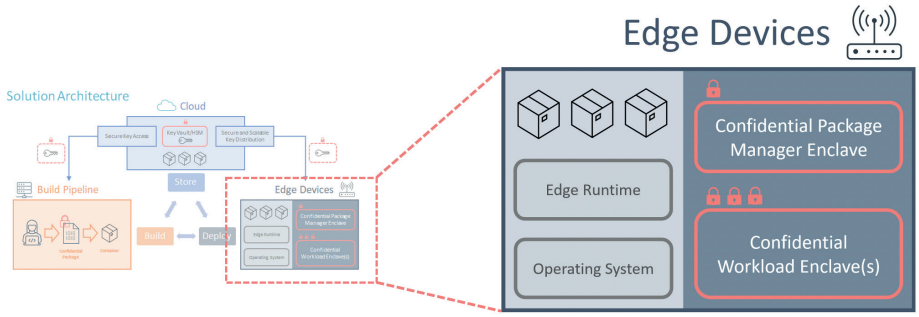
The key management architecture is central to ensuring end to end security. The architecture leverages the sharing of a common symmetrical encryption key between the build pipeline and the target device. This way, the build pipeline can produce an encrypted application that only the target device can decrypt an execute. This key is protected at all times by storing it securely in a key Vault or managed HSM in the cloud. Whenever the key is shared, it is protected with additional encryption, eg. Edge Device public key. The concept of a one-time use key for transport is referred to as a class key.

Build Pipeline



The application development is simplified by leveraging *OpenEnclave SDK* and Visual Studio IDE to create a uniform and familiar development environment. The application can be debugged and tested locally using QEMU. It can then be encrypted with the symmetrical key and embedded in a confidential package to prepare it for secure storage and transport. Finally, the application is distributed in a container where it sits securely in the confidential package.

Edge Devices



The device needs to have both the Trusted Execution environment: OPTEE and the Confidential Package Manager running on it. We also need the IOT Edge runtime and a container runtime. The Host OS is less critical. The OS used for real deployment can differ, but a hardened Yocto build is recommended, or the use of a binary Linux release tailored to long time support.

Real world realization of the Enclave Device Blueprint

This real-world realization of the Enclave Device blueprint is available to help you start your journey in confidential computing at the edge.

The Scalys TrustBox Edge 201 comes with hardware backed security and software attestation. The box is in the process of meeting the Arm SystemReady certification requirements and is ready to support the confidential packaging system from Azure Cloud interface. This provides a great platform to take the blueprint and create a customer ready solution.

With a Scalys TrustBox Edge 201 and the opensource software, you have everything you need to get started. At the Project home page, you will find information on the enclave device, key components, solution, and demo.

Just The Beginning

We view the IoT as a community that benefits from collaboration, and we think IoT deployments succeed when trust spans from the device manufacturer to the solution integration and the solution backend. Confidential computing has an important role to play in that chain of trust, by protecting data when it's in use, either in the cloud or at the edge.

The model we present here for developing a secure enclave IoT device is just one example of how the migration of confidential computing to the IoT can be made easier. As the IoT industry works together to expand the ecosystem for secure applications, developers will be able to access a wider variety of building blocks.

Our work in trusted devices, including our support for the approach described here, is a natural extension of our efforts to simplify development on a broader scale, from device to cloud.

Learn more about the Enclave Device Blueprint from the project page.

Whether it's core projects, such as the SystemReady compliance certification program for hardware and firmware standards, or the Microsoft Azure IoT cloud services for cloud-native development, we are committed to make it easier to develop and deploy secure solutions.

The ultimate goal of all these efforts is to abstract the complexities away from solution builds, so developers can spend less time developing and debugging baseline security features, and spend more time focused on new ideas that add value.

Join Us

The effort to migrate confidential computing to the IoT benefits from your involvement. We invite you to learn more about confidential computing and how it relates to the IoT, and encourage you to join us as part of the broader community that's working to build the ecosystem for confidential IoT development.

The CCC website (https://confidentialcomputing.io), which includes links to white papers and open-source projects, is a good place to start.

Also join the projects that provide the building blocks to facilitate delivery, management and monitoring of applications in confidential compute environments at the edge and in IoT:

Arm Project Cassini	An initiative to help deliver a cloud-native software experience across a secure Arm edge ecosystem
Parsec	To help abstract RoT and support key handling functions.
OpenEnclave SDK	To help with the development of applications that leverage TEEs
Arm SystemReady	To ensure standard based hardware and firmware.
Trustbox Edge	Connect with Scalys to accelerate your development and deployments of secure IoT devices.