What Is Confidential Computing?

Confidential computing is a way of processing data in a protected zone of a computer's processor, often inside a remote edge or public cloud server, and proving that no one viewed or altered the work.

Author: Rick Merritt

Cloud and edge networks are setting up a new line of defense, called confidential computing, to protect the growing wealth of data users process in those environments.

Confidential computing is a way of protecting data in use, for example while in memory or during computation, and preventing anyone from viewing or altering the work.

Using cryptographic keys linked to the processors, confidential computing creates a trusted execution environment or secure enclave. That safe digital space supports a cryptographically signed proof, called attestation, that the hardware and firmware is correctly configured to prevent the viewing or alteration of their data or application code.

In the language of security specialists, confidential computing provides assurances of data and code privacy as well as data and code integrity.

Confidential computing is a relatively new capability for protecting data in use.

For many years, computers have used encryption to protect data that's in transit on a network and data at rest, stored in a drive or non-volatile memory chip. But with no practical way to run calculations on encrypted data, users faced a risk of having their data seen, scrambled or stolen while it was in use inside a processor or main memory.

With confidential computing, systems can now cover all three legs of the data-lifecycle stool, so data is never in the clear.

In the past, computer security mainly focused on protecting data on systems users owned, like their enterprise servers. In this scenario, it's okay that system software sees the user's data and code.

With the advent of cloud and edge computing, users now routinely run their workloads on computers they don't own. So confidential computing flips the focus to protecting the users' data from whoever owns the machine.

With confidential computing, software running on the cloud or edge computer, like an operating system or hypervisor, still manages work. For example, it allocates memory to the user program, but it can never read or alter the data in memory allocated by the user.

A 2015 research paper was one of several using new Security Guard Extensions (Intel SGX) in x86 CPUs to show what's possible. It called its approach VC3, for Verifiable Confidential Cloud Computing , and the name — or at least part of it — stuck.

"We started calling it confidential cloud computing," said Felix Schuster, lead author on the 2015 paper.

Four years later, Schuster co-founded Edgeless Systems, a company in Bochum, Germany, that develops tools so users can create their own confidential-computing apps to improve data protection.

Confidential computing is "like attaching a contract to your data that only allows certain things to be done with it," he said.

Taking a deeper look, confidential computing sits on a foundation called a root of trust, which is based on a secured key unique to each processor.

The processor checks it has the right firmware to start operating with what's called a secure, measured boot. That process spawns reference data, verifying the chip is in a known safe state to start work.

Next, the processor establishes a secure enclave or trusted execution environment (TEE) sealed off from the rest of the system where the user's application runs. The app brings encrypted data into the TEE, decrypts it, runs the user's program, encrypts the result and sends it off.

At no time could the machine owner view the user's code or data.

One other piece is crucial: It proves to the user no one could tamper with the data or software.

The proof is delivered through a multi-step process called attestation (see diagram above).

The good news is researchers and commercially available services have demonstrated confidential computing works, often providing data security without significantly impacting performance.

As a result, users no longer need to trust all the software and systems administrators in separate cloud and edge companies at remote locations.

Confidential computing closes many doors hackers like to use. It isolates programs and their data from attacks that could come from firmware, operating systems, hypervisors, virtual machines — even physical interfaces like a USB port or PCI Express connector on the computer.

The new level of security promises to reduce data breaches that rose from 662 in 2010 to more than 1,000 by 2021 in the U.S. alone, according to a report from the Identity Theft Resource Center .

That said, no security measure is a panacea, but confidential computing is a great security tool, placing control directly in the hands of "data owners".

Users with sensitive datasets and regulated industries like banks, healthcare providers and governments are among the first to use confidential computing. But that's just the start.

Because it protects sensitive data and intellectual property, confidential computing will let groups feel they can collaborate safely. They share an attested proof their content and code was secured.

Example applications for confidential computing include:

Companies executing smart contracts with blockchains

Research hospitals collaborating to train AI models that analyze trends in patient data

Retailers, telecom providers and others at the network's edge, protecting personal information in locations where physical access to the computer is possible

Software vendors can distribute products which include AI models and proprietary algorithms while preserving their intellectual property

While confidential computing is getting its start in public cloud services, it will spread rapidly.

Users need confidential computing to protect edge servers in unattended or hard-to-reach locations. Enterprise data centers can use it to guard against insider attacks and protect one confidential workload from another.

So far, most users are in a proof-of-concept stage with hopes of putting workloads into production soon, said Schuster.

Looking forward, confidential computing will not be limited to special-purpose or sensitive workloads. It will be used broadly, like the cloud services hosting this new level of security.

Indeed, experts predict confidential computing will become as widely used as encryption.

The technology's potential motivated vendors in 2019 to launch the Confidential Computing Consortium , part of the Linux Foundation. CCC's members include processor and cloud leaders as well as dozens of software companies.

The group's projects include the Open Enclave SDK , a framework for building trusted execution environments.

“Our biggest mandate is supporting all the open-source projects that are foundational parts of the ecosystem,” said Jethro Beekman, a member of the CCC’s technical advisory council and vice president of technology at Fortanix, one of the first startups founded to develop confidential computing software.

“It’s a compelling paradigm to put security at the data level, rather than worry about the details of the infrastructure — that should result in not needing to read about data breaches in the paper every day,” said Beekman, who wrote his 2016 Ph.D. dissertation on confidential computing.

Implementations of confidential computing are evolving rapidly.

At the CPU level, AMD has released Secure Encrypted Virtualization with Secure Nested Paging (SEV-SNP). It extends the process-level protection in Intel SGX to full virtual machines, so users can implement confidential computing without needing to rewrite their applications.

Top processor makers have aligned on supporting this approach. Intel’s support comes via new Trusted Domain Extensions . Arm has described its implementation, called Realms .

Proponents of the RISC-V processor architecture are implementing confidential computing in an open-source project called Keystone .

NVIDIA is bringing GPU acceleration to VM-style confidential computing to market with its Hopper architecture GPUs .

The H100 Tensor Core GPUs enable confidential computing for a broad swath of AI and high performance computing use cases. This gives users of these security services access to accelerated computing .

Meanwhile, cloud service providers are offering services today based on one or more of the underlying technologies or their own unique hybrids.

Over time, industry guidelines and standards will emerge and evolve for aspects of confidential computing such as attestation and efficient, secure I/O, said Beekman of CCC.

While it’s a relatively new privacy tool, confidential computing’s ability to protect code and data and provide guarantees of confidentiality makes it a powerful one.

Looking ahead, experts expect confidential computing will be blended with other privacy methods like fully homomorphic encryption (FHE), federated learning, differential privacy, and other forms of multiparty computing.

Using all the elements of the modern privacy toolbox will be key to success as demand for AI and privacy grows.

So, there are many moves ahead in the great chess game of security to overcome the challenges and realize the benefits of confidential computing.

To learn more, watch “Hopper Confidential Computing: How it Works Under the Hood,” session S51709 at GTC on March 22 or later (free with registration).

Check out “Confidential Computing: The Developer’s View to Secure an Application and Data on NVIDIA H100,” session S51684 on March 23 or later.

You also can attend a March 15 panel discussion at the Open Confidential Computing Conference moderated by Schuster and featuring Ian Buck, NVIDIA’s vice president of hyperscale and HPC. In addition, Mark Overby, NVIDIA’s chief platform security architect, will host a session there on “ Attesting NVIDIA GPUs in a Confidential Computing Environment .”

And watch the video below.

Original URL: https://blogs.nvidia.com/blog/2023/03/01/what-is-confidential-computing/