Confidential Computing in the Cloud

Viet Anh Pham
Bachelor of Science (Honours) in Cybercrime and IT Security
Southeast Technological University (SETU), Carlow Campus
Portlaoise, Co. Laois
Yan2020pham@gmail.com

***Abstract*—Confidential Computing is an emerging paradigm in cloud security that protects data during active use — a phase traditionally unprotected by conventional mechanisms. By leveraging hardware-based Trusted Execution Environments (TEEs) such as Intel SGX, AMD SEV-SNP, and ARM TrustZone, it enables secure execution of sensitive workloads in isolated environments. This report provides a comprehensive overview of the Confidential Computing landscape, covering its architecture, practical use cases across finance, healthcare, and AI, and the technical challenges it faces, including performance overhead and lack of standardization. It also explores the future potential of this technology in enhancing privacy, regulatory compliance, and multi-party collaboration across cloud platforms.**

Keywords— Confidential Computing, Trusted Execution Environments (TEEs), Intel SGX, AMD SEV, ARM TrustZone, Cloud Security, Data in Use, Remote Attestation, Secure AI, Privacy-Preserving Computing

I. Introduction

Cloud computing has changed how organizations build and run digital systems. It provides flexible, scalable, and on-demand access to computing resources, which makes it ideal for modern applications. However, because cloud environments are shared and virtualized, there are growing concerns about how to keep sensitive data secure, especially when that data is actively being used. Most cloud security today focuses on protecting data when it is stored, known as "data at rest," or when it is transmitted over networks, referred to as "data in transit." While these methods are effective, they do not protect data while it is being processed in memory. This phase, known as "data in use," stays at a weak point in cloud security. During this time, even privileged users such as cloud administrators or attackers with access to the underlying system could potentially view the data.

Confidential Computing [1] addresses this issue by using hardware-based technology called Trusted Execution Environments [2] (TEEs). TEEs create a secure, isolated space within the CPU where data can be processed while being still protected from the rest of the system. Technologies like Intel SGX [3], AMD SEV-SNP, and ARM Trust Zone [7] make this possible by enabling data confidentiality and integrity even when running in public cloud environments. Major cloud providers, including AWS, Azure, and Google Cloud [8], have started integrating this approach into their infrastructure to support sensitive workloads such as financial services, healthcare processing, and AI inference.

Initially, I planned to research how artificial intelligence is used in the cloud. However, after doing some reading, I found the topic less engaging than I expected. During that process, I discovered Confidential Computing [1] and found it far more relevant to my interests in cybersecurity. It combines technical innovation with practical applications and addresses a significant gap in how we secure cloud-based systems. This report will explore the architecture of Confidential Computing [1], how it is currently used in cloud environments, the challenges involved, and potential future directions.

II. Literature review

Cloud security has traditionally focused on using encryption and access control to safeguard data while it's being transferred or stored. But the idea of protecting data while it's being actively used, referred to as securing 'data in use' has only recently started to get significant attention. To bridge this gap, the concept of Confidential Computing [1] appeared, using hardware-based isolation to provide stronger protections. At the heart of this approach are Trusted Execution Environments (TEEs) [2], which are secure, isolated memory areas within a CPU designed to keep data and code safe from unauthorized access.

One of the first and most well-known examples of a TEE is Intel’s Software Guard Extensions (SGX). SGX lets applications create secure memory areas called 'enclaves,' which are protected from everything else on the system, including the operating system and hypervisor. Though initially intended for client-side computing, SGX helped lay the foundation for bringing Confidential Computing [1] to the cloud. Other key technologies include AMD’s Secure Encrypted Virtualization (SEV) and its extension SEV-SNP, which protects entire virtual machines through memory encryption and integrity checks. ARM TrustZone [7], another TEE technology, is often used in mobile and embedded systems and plays a role in advancing secure hardware platforms overall.

Studies have shown that TEEs can effectively protect sensitive data in the cloud. Research on topics like secure multi-party computation, private machine learning, and protected key management illustrates how TEEs support the safe handling of confidential tasks. Still, TEEs aren't without flaws. Academic studies have revealed that side-channel attacks where an attacker analyzes things like access patterns, power usage, or timing can sometimes reveal sensitive information. These findings have sparked ongoing research into making TEEs more secure and integrating them with other protective measures.

Considering these trends, major cloud providers have begun adding TEE support to their services. Microsoft Azure [6] launched Confidential VMs based on AMD SEV, while Google Cloud [8] introduced similar functionality. AWS took a unique approach with its Nitro Enclaves [5], which use the Nitro hypervisor to create isolated environments within EC2 instances for handling sensitive tasks. Alongside these efforts, open-source projects like Confidential Containers [11] (CoCo), backed by the Cloud Native Computing Foundation (CNCF), enabling secure processing in containerized setups. These initiatives show that Confidential Computing [1] is gaining traction and may soon become a core part of cloud security strategies.

In summary, while Confidential Computing [1] is still evolving, it's clear that it plays a vital role in modern cloud security. Building on solid research in cryptography and hardware isolation, it also opens new possibilities for secure software development in the cloud, trusted attestation methods, and privacy-focused computing.

III. System Architecture

Confidential Computing [1] uses Trusted Execution Environments (TEEs) [2] to safeguard data while it's being actively processed. A TEE is a secure, hardware-isolated space inside the CPU that keeps both the code and data safe from tampering or exposure even from system-level software like the operating system, hypervisor, or cloud provider.

A. Trusted Execution Environments (TEEs)

Each hardware vendor has developed TEEs in ways tailored to several types of workloads:

- **Intel SGX [3] (Software Guard Extensions)**: This lets applications define secure enclaves that shield selected parts of their code and data. It provides fine-grained isolation but is limited by memory capacity and often requires changing the app's code.
- **AMD SEV-SNP**: This technology encrypts the memory of full virtual machines and ensures their integrity, making it well-suited for cloud scenarios where users want strong isolation without altering their applications.
- **ARM TrustZone [7]**: It splits devices into two execution zones: a "secure world" and a "normal world." It's widely adopted in mobile and embedded systems and is being explored for edge computing use cases.

In cloud deployments, TEEs are often integrated with VMs or containers, allowing sensitive apps to run securely without needing to trust the host infrastructure.

B. Confidential Computing Workflow

A typical Confidential Computing [1] process in the cloud includes these stages:

- **Encrypted Application Packaging**: The app is packaged as a VM or container and encrypted before deployment.
- **TEE Initialization**: The app launches in a secure, TEE-enabled environment like an Azure Confidential VM or AWS Nitro Enclave.
- **Remote Attestation [2]**: The TEE generates a cryptographic attestation report to prove it's running genuine, unmodified code.
- **Key Provisioning**: This report is sent to a Key Management Service (KMS). If validated, the KMS sends the necessary decryption keys to the enclave.
- **Secure Execution**: Inside the TEE, the app decrypts and processes data without exposing it outside the secure zone.
- **(Optional) Secure Output**: Processed data can be encrypted again before it's sent to external storage or systems.

This approach ensures data is protected through every phase including during processing, which is typically the most vulnerable step in cloud computing.

C. Remote Attestation

Remote attestation is a critical mechanism that enables trust in Confidential Computing [1] environments. It allows a relying party such as a user, external service, or key management system to verify that a workload is running inside a genuine Trusted Execution Environment (TEE) on trusted hardware, and that the software loaded into the TEE has not been tampered with.

When a TEE is initialized, it generates a cryptographic report that includes a measurement of the system's current state, including the enclave's code and configuration. This report was signed using a key rooted in the hardware manufacturer, such as Intel or AMD. The signed attestation report can then be sent to a remote party, which verifies its authenticity using a trusted attestation service.

If the attestation is successful, the remote party typically a Key Management Service (KMS) releases sensitive assets such as encryption keys or credentials. These assets are securely delivered to the enclave, where they are used to decrypt and process confidential data. This ensures that secrets are only made available to workloads that are probably running in a secure and trusted environment.

Remote attestation is supported by all major cloud Confidential Computing [1] offerings. For example, AWS Nitro Enclaves [5] use attestation to interact securely with AWS KMS, while Azure Confidential VMs provide attestation APIs for verifying virtual machine integrity. This process forms the backbone of trust in Confidential Computing [1] architectures, particularly in public cloud environments where users do not control the physical hardware.

D. Cloud Provider Implementations

Top cloud platforms support Confidential Computing [1] using various strategies:

- Microsoft Azure [6]: Offers Confidential VMs using AMD SEV-SNP, which protect workloads without requiring app changes. Azure also supports Intel SGX [3] for use cases needing enclave-level customization, like secure databases or analytics.
- Amazon Web Services (AWS) [4]: Provides Nitro Enclaves, which create secure, stripped-down computer environments within EC2 instances. They lack network or storage access by default, minimizing risk. Built-in attestation works with AWS KMS.
- Google Cloud [8]: Supports Confidential VMs on its Compute Engine platform using AMD SEV. These VMs encrypt memory and CPU state, making it easy to run standard applications securely, without needing code changes.
- Confidential Containers [11] (CoCo): An open-source CNCF project that allows Kubernetes workloads to run in TEEs via Kata Containers. It brings Confidential Computing [1] to containerized, cloud-native applications without requiring changes to the apps.

E. Architecture Components Summary

A Confidential Computing [1] system typically includes the following components:

- User/DevOps Tools: Used to package, encrypt, and deploy the application or container image.
- TEE-enabled Cloud Instance: The secure runtime environment where workloads are isolated from the host system.
- Remote Attestation [2] Service: Verifies the integrity and authenticity of the TEE by validating cryptographic measurements.
- Key Management System (KMS): Holds cryptographic keys and only releases them after successful attestation.
- Execution Environment: The enclave or secure VM where decrypted data is processed safely and confidentially.

Together, these components form a trusted execution pipeline that enhances data privacy and security in public cloud environments, especially for sensitive workloads.

IV. Use cases

Confidential Computing [1] enables organizations to process sensitive data securely in public cloud environments, where trust in the underlying infrastructure is limited. It is especially valuable in industries with strict privacy, security, or compliance requirements. The following use cases show how Confidential Computing [1] can be applied in real-world scenarios.

A. 4.1 Secure Machine Learning (Confidential AI)

Machine learning models often rely on extremely sensitive data like health records, financial transactions, or behavioral patterns. Running AI workloads on public cloud platforms can expose this data to risks if not properly protected.

Confidential Computing [1] enables the entire AI pipeline from data cleaning to model training and inference to run within a Trusted Execution Environment (TEE). This setup ensures that:

- Data stays encrypted until it's securely processed inside the TEE.
- Proprietary models and algorithms are shielded from theft or tampering.
- Input and output data are still hidden from the cloud provider and host system.

Cloud platforms such as Azure Confidential ML and Intel's Confidential AI already offer support for secure AI execution, making this approach suitable for sectors like finance, defense, and healthcare.

B. 4.2 Financial Services and Secure Transactions

The financial sector deals with extremely sensitive information, including personal identifiers, credit histories, and transaction records. Regulations like PCI DSS and GDPR require that such data is protected during all stages of processing.

Confidential Computing [1] allows financial institutions to:

- Perform secure credit scoring, fraud detection, or transaction analysis in cloud environments.
- Run workloads in isolated enclaves to protect against insider threats.
- Ensure compliance by processing data only within verified environments using remote attestation.

Banks and fintech companies are exploring Confidential VMs and Nitro Enclaves to meet these requirements without building and managing their own data centers.

C. 4.3 Healthcare and Privacy-Preserving Analytics

Healthcare data is among the most regulated and sensitive information handled in the cloud. Whether analyzing medical imaging, patient records, or genomic data, organizations must guarantee patient privacy.

Using Confidential Computing [1], hospitals, researchers, and insurers can:

- Run analytics on encrypted health records without exposing patient data.
- Enable cross-institutional collaboration (e.g., research using confidential data from multiple hospitals) without sharing raw datasets.
- Remain compliant with HIPAA, GDPR, and other privacy regulations.

This approach also supports privacy-preserving technologies like federated learning, where multiple organizations train AI models without transferring or revealing local data.

D. 4.4 Multi-Party Computation and Data Collaboration

Many organizations face challenges when collaborating on sensitive data — such as joint research, fraud detection across banks, or supply chain auditing. Traditional models require full data sharing, which may not be legally or commercially practical.

Confidential Computing [1] provides a solution by allowing multiple parties to:

- Upload encrypted data to a shared TEE in the cloud.
- Perform joint computation inside the enclave without revealing their own datasets.
- Trust that the results are correct and confidential, thanks to remote attestation and hardware-backed integrity guarantees.

This use case is particularly relevant for government collaborations, cross-border data analysis, and enterprise partnerships that demand both privacy and mutual trust.

E. 4.5 Secure Edge and IoT Data Processing

In edge and IoT environments, data is collected and analyzed locally on devices like sensors, gateways, or autonomous machines. This information, whether personal or industrial, must be protected, especially in untrusted or remote settings.

Confidential Computing [1] can now be applied at the edge, thanks to technologies like ARM TrustZone [7] and

newer enclave-enabled processors. These allow secure processing directly on low-power, embedded systems.

With this setup, organizations can:

- Process data locally on edge devices while keeping it hidden from the host.
- Make private, secure decisions (e.g., anomaly detection, control logic) without exposing sensitive logic.
- Send only encrypted or anonymized data to the cloud saving bandwidth and reducing risk.

Relevant applications include smart city tech, autonomous vehicles, industrial systems, and real-time healthcare monitoring where low latency and data confidentiality are equally important.

V. Limitations and Challenges

Confidential Computing [1] introduces strong protections for data in use, but it also brings several limitations and challenges that affect its performance, usability, and adoption. These challenges are important to understand when evaluating whether and how to integrate Confidential Computing [1] into real-world systems.

A. Hardware Limitations

One of the main technical hurdles with Trusted Execution Environments (TEEs) [2] is limited memory. For instance, early versions of Intel SGX [3] capped enclave memory at around 128 MB, which made it tough to support large applications without relying on paging, a process that significantly slows performance. Although newer technologies like AMD SEV-SNP and Intel TDX offer better scalability, they still face challenges when running memory-heavy workloads, such as big machine learning models or in-memory databases.

B. Performance Overhead

Running workloads in secure enclaves introduces performance overhead due to isolated memory access, encryption/decryption operations, and remote attestation procedures. These overheads may result in increased latency or reduced throughput, especially for real-time or high-performance applications. Although more recent TEEs offer better hardware acceleration and optimizations, developers must still consider the trade-off between enhanced security and reduced system performance.

C. Side-Channel Vulnerabilities

TEEs are not immune to side-channel attacks, which exploit indirect information such as timing, cache access patterns, or speculative execution behavior. Well-known vulnerabilities like Spectre, Foreshadow, and ZombieLoad have proven that even isolated environments can be compromised through sophisticated techniques. While mitigations exist and vendors continue to enhance protections, these vulnerabilities highlight the need for continuous monitoring, patching, and secure enclave-aware programming practices.

D. Development and Debugging Complexity

Writing software for TEEs isn't straightforward. Developers often need to use vendor-specific SDKs, work within tight memory and instruction limits, and refactor applications to meet TEE constraints. On top of that, debugging is difficult because standard monitoring tools can't access secure parts of memory. This makes development slower and increases the learning curve for teams adopting this technology.

E. Lack of Standardization

The current Confidential Computing ecosystem is fragmented across vendors and platforms. Each major cloud provider AWS, Azure, Google Cloud [8] uses different hardware backends, APIs, and attestation mechanisms. For example, AWS Nitro Enclaves [5] differ significantly from Azure SGX-based VMs, or Google Cloud [8]'s SEV-based Confidential VMs. This lack of standardization creates challenges in developing portable applications, implementing multi-cloud architectures, and integrating third-party tools. Efforts by organizations such as the Confidential Computing Consortium [1] (CCC) aim to set up common frameworks and interfaces, but full cross-platform interoperability is still a work in progress.

F. Cost Implications

TEEs are usually tied to premium cloud instances, which can drive up costs. Additionally, the complexity involved in securely designing, deploying, and maintaining these workloads adds operational overhead. For organizations with tight budgets or limited technical resources, these costs might outweigh the security benefits, especially if alternatives already meet their compliance or privacy needs.

VI. Discussion and Future Directions

Confidential Computing [1] marks a major step forward in cloud security, filling a long-standing gap by securing data during processing. By isolating workloads inside Trusted Execution Environments (TEEs) [2], organizations can confidently handle sensitive information even in multi-tenant or less trusted cloud settings. As interest in technology grows, several trends are shaping its evolution and broader adoption.

A. Discussion

One significant development is the expansion of Confidential Computing [1] beyond traditional virtual machines. It's now being extended to containers, managed by Kubernetes services, and hardware accelerators. For instance, in January 2025, Google Cloud [8] launched Confidential GKE Nodes [9] on its C3D series using AMD's 4th Gen EPYC ("Genoa") chips. These nodes use AMD SEV to encrypt data in use, letting Kubernetes workloads benefit from strong security without any need for code changes. This shows a clear move toward making Confidential Computing [1] more cloud-native and developer-friendly.

Open-source projects are also helping drive adoption. Tools like Confidential Containers [11] (CoCo), part of the Cloud Native Computing Foundation (CNCF), make it easier to run containerized apps inside secure, TEE-backed virtual machines. By integrating this tech into Kubernetes and CI/CD pipelines, they're helping developers build secure systems with less friction.

On the hardware side, newer TEEs like Intel TDX and AMD SEV-SNP are improving scalability, memory capacity, and security compared to older solutions like Intel SGX [3]. These enhancements reduce performance trade-offs and expand the range of applications that can run securely inside enclaves.

However, the ecosystem is still fragmented. Each cloud provider has its own approach, with unique APIs, attestation processes, and toolchains. This lack of standardization makes it hard to move applications across platforms or build seamless multi-cloud environments. While the Confidential Computing Consortium [1] is working to unify standards, full interoperability is still a work in progress.

Other challenges include limited development tools, debugging constraints, and the need to meet complex compliance requirements. Addressing these issues will be crucial for broader adoption, especially in industries that handle sensitive or regulated data.

B. Future Directions

Looking ahead, Confidential Computing [1] is poised to become a core part of secure cloud architecture. As cloud-native development continues to grow, there is a strong movement toward seamless integration of TEEs into containers, serverless functions, and managed services all without requiring developers to change application code.

The rise of open-source projects like CoCo is expected to accelerate this shift, enabling enclave-backed containers to be deployed and orchestrated alongside standard workloads. This paves the way for broader adoption in DevSecOps pipelines and multi-cloud environments.

At the application level, Confidential Computing [1] is opening the door to novel privacy-preserving workloads, including:

- Federated learning and collaborative AI, where institutions share model insights without exposing raw data.
- Confidential blockchain computation, where smart contracts can be executed privately within enclaves.
- Secure edge computing, where sensitive data is processed locally on IoT or embedded devices using technologies like ARM TrustZone [7].

The industry is also likely to see progress in cross-cloud attestation, where workloads spanning multiple cloud providers can verify trust relationships and securely exchange data. In parallel, regulatory bodies may begin to formalize standards for enclave certification and auditability, helping organizations meet compliance requirements more easily.

Ultimately, as hardware becomes more capable, tooling improves, and standards mature, Confidential Computing [1] is expected to become a default option for securing high-value workloads. It has the potential to fundamentally reshape how privacy and trust are enforced in modern cloud infrastructures.

VII. Conclusion

As more industries move their operations to the cloud, protecting the confidentiality and integrity of sensitive data has never been more critical. Traditional security measures typically focus on keeping data safe when it's stored or being transmitted, but they fall short when it comes to protecting data while it's actively being used. Confidential Computing [1] fills this gap by securing data during processing when it's often most vulnerable.

By using Trusted Execution Environments (TEEs) [2], Confidential Computing [1] provides secure, hardware-based isolation for workloads, even in multi-tenant or untrusted environments. What was once a niche concept was quickly gained traction, evolving into a suite of cloud-native services from major providers like AWS, Microsoft Azure [6], and Google Cloud [8]. These platforms support a wide range of valuable use cases, including secure machine learning, privacy-preserving analytics, collaborative computation, and edge data processing.

That said, technology isn't without its hurdles. Current limitations include constrained TEE resources, performance slowdowns, complex development requirements, vulnerability to side-channel attacks, and inconsistent implementations across platforms. However, the landscape is improving thanks to coordinated industry efforts. Organizations such as the Confidential Computing Consortium [1] (CCC), the release of more advanced hardware like Intel TDX and AMD SEV-SNP, and open-source tools like Confidential Containers [11] are helping overcome these challenges.

Looking to the future, Confidential Computing [1] is on track to become a foundational element of secure cloud infrastructures. As it becomes more tightly integrated with containerized workflows, edge environments, and serverless architectures, developers will find it easier to adopt without needing to change existing code.

There are also exciting opportunities for research and innovation, such as enhancing remote attestation protocols, defending against new forms of side-channel attacks, and developing systems that enable secure workload orchestration across multiple cloud platforms. On a practical level, organizations can start exploring this technology by testing secure AI applications, enabling confidential data sharing between institutions, or processing sensitive edge data using enclave-ready devices.

Ultimately, Confidential Computing [1] marks a shift toward embedding trust into the very infrastructure of cloud computing. It empowers users to control their data more effectively, supports compliance with strict regulations, and opens the door to entirely new types of applications focused on privacy and security.

References

[1] Confidential Computing Consortium, "About Confidential Computing," *confidentialcomputing.io*, 2024. [Online]. Available: https://confidentialcomputing.io/

[2] Confidential Computing Consortium, "Basics of Trusted Execution Environments (TEEs): The Heart of Confidential Computing," *confidentialcomputing.io*, Mar. 2024. [Online]. Available: https://confidentialcomputing.io/2024/03/13/basics-of-trusted-execution-environments-tees-the-heart-of-confidential-computing/

[3] J. Intel Corporation, "Intel® Software Guard Extensions (Intel® SGX),"[Online].Available: https://www.intel.com/content/www/us/en/architecture-and-technology/software-guard-extensions.html

[4] Amazon Web Services (AWS), "AWS Nitro Enclaves," *AWS Documentation*,2024.[Online].Available: https://docs.aws.amazon.com/enclaves/latest/user/nitro-enclave.html

[5] Amazon Web Services, "AWS Nitro Enclaves," [Online]. Available: https://aws.amazon.com/ec2/nitro/nitro-enclaves/

[6] Microsoft Azure, "Azure Confidential Computing," *Microsoft Learn*, 2024. [Online]. Available: https://learn.microsoft.com/en-us/azure/confidential-computing/

[7] ARM Ltd., "ARM TrustZone technology for the ARMv8-M architecture Version 2.0," *ARM Developer Documentation*, 2023. [Online].Available:

https://developer.arm.com/documentation/100690/0200/ARM-TrustZone-technology

[8] Google Cloud, "Confidential Computing," *Google Cloud Documentation*,2024.[Online].Available: https://cloud.google.com/confidential-computing

[9] J. Young and R. Kolga, "Privacy-preserving Confidential Computing now on even more machines and services," *Google Cloud Blog*, Jan. 28,2025.[Online].Available: https://cloud.google.com/blog/products/identity-security/privacy-preserving-confidential-computing-now-on-even-more-machines.

[10] NVIDIA, "Confidential Computing Solutions," *NVIDIA Data Center Solutions*, 2024. [Online]. Available: https://www.nvidia.com/en-us/data-center/solutions/confidential-computing/

[11] Kubernetes Blog, "Confidential Kubernetes: A Secure Approach for Sensitive Workloads," *Kubernetes.io*, Jul. 6, 2023. [Online]. Available: https://kubernetes.io/blog/2023/07/06/confidential-kubernetes/