Elastic Cloud Server

QingTian System Security Technical White Paper

Issue 01

Date 2025-10-17

HUAWEI

HUAWEI CLOUD COMPUTING TECHNOLOGIES CO., LTD.

Trademarks and Permissions


and other Huawei trademarks are the property of Huawei Technologies Co., Ltd.
All other trademarks and trade names mentioned in this document are the property of their respective holders.

Notice

The purchased products, services and features are stipulated by the contract made between Huawei Cloud and the customer. All or part of the products, services and features described in this document may not be within the purchase scope or the usage scope. Unless otherwise specified in the contract, all statements, information, and recommendations in this document are provided "AS IS" without warranties, guarantees or representations of any kind, either express or implied.

The information in this document is subject to change without notice. Every effort has been made in the preparation of this document to ensure accuracy of the contents, but all statements, information, and recommendations in this document do not constitute a warranty of any kind, express or implied.

Huawei Cloud Computing Technologies Co., Ltd.

Address: Huawei Cloud Data Center Jiaoxinggong Road
Qianzhong Avenue
Gui'an New District
Gui Zhou 550029
People's Republic of China

Website: https://www.huaweicloud.com/intl/en-us/

Contents

1 QingTian System Security Technical White Paper

1.1 About This Document

Huawei Cloud Elastic Cloud Server (ECS) provides scalable cloud servers that offer secure, reliable, and high-performance compute resources. The QingTian system is the underlying virtualization platform for QingTian ECSs. It is a combination of custom-designed servers, data processors, system management components, and dedicated firmware for cloud data centers.

The QingTian system architecture supports multiple forms (VMs, bare metal servers, and containers) and heterogeneous compute. Based on the QingTian system, Huawei Cloud builds infrastructure cloud services with higher security, isolation, performance, and lower costs. QingTian also provides trusted computing, confidential computing, and a series of security features for multi-tenant isolation and cloud service isolation.

Huawei Cloud's top priority is to ensure the confidentiality, integrity, and availability of customers' workloads. It continuously invests in key security technologies and engineering best practices to meet and even exceed the most demanding customers' requirements for data security and privacy protection on the cloud.

This document describes the security design of the QingTian system and the multi-dimensional isolation capabilities provided based on the QingTian system to help you evaluate the applicability of ECS to sensitive workloads.

- **QingTian System Overview**: describes virtualization technologies and the architecture changes after QingTian is introduced.
- **QingTian Threat Assumptions and Security Methods**: describes the threat assumptions and security design methods of the QingTian system.
- **QingTian System Components**: describes the key security design of QingTian system components, including QingTian Cards (QingTian Controller and I/O offloading cards) and QingTian Hypervisor.
- **QingTian Confidential Computing**: describes the design concepts of QingTian confidential computing, including isolation design in two dimensions and cryptographic attestation.

- **From Physical Isolation to Logical Isolation**: describes how to enhance a series of security isolation technologies from physical isolation to logical isolation based on QingTian.
- **Zero-Privilege O&M**: describes the zero-privilege O&M concept of Huawei Cloud production systems and key security system protection practices.
- **Case: Secure Cloud Migration of Financial Customer Data**: provides a design reference for financial customers to protect cloud data security.
- **Conclusion**: summarizes the functions and advantages of the QingTian system.

1.2 QingTian System Overview

Traditional Virtualization System

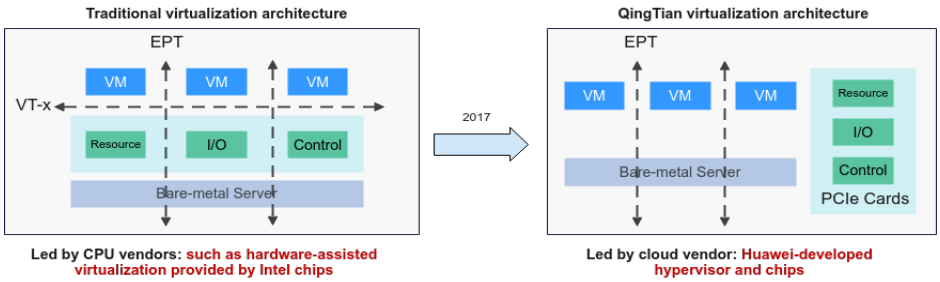
Virtualization enables multiple operating systems (OSs) to run on a single physical computer system. A virtualization system leverages hardware-assisted virtualization to virtualize CPUs, memory, and I/Os, and provides independent, isolated hardware space for customers' virtual OSs. Virtualization can divide the resources of a single server among multiple tenant VMs. Each VM is allocated isolated resources, thereby effectively utilizing the resources of a single server.

The core component of a virtualization system is the hypervisor. The core responsibilities of the hypervisor include abstracting and allocating physical hardware resources, managing the full lifecycle of VMs, and isolating VMs. The hypervisor can isolate resources of different VMs on the same physical server to prevent data theft or malicious attacks between VMs and ensure that the VM resource usage is not affected by peripheral VMs. VMs are isolated for security. End users can only access resources (such as hardware and software resources and data) allocated to their own VMs.

QingTian Virtualization System Evolution

The QingTian system architecture is the next generation of hardware-software synergy architecture launched by Huawei Cloud. It provides key capabilities such as zero resource reservation, zero compute loss, zero service jitter, and strong security isolation. Huawei Cloud has released QingTian instances in 2017. The QingTian architecture supports multiple forms and heterogeneous compute, such as VMs, bare metal servers, and containers. After years of development, the QingTian system has reshaped Huawei Cloud infrastructure and has become the mainstream underlying platform of the new generation of instances.

Figure 1-1 System evolution



The QingTian system consists of Huawei-developed QingTian Cards and QingTian Hypervisor.

- QingTian Cards are Huawei Cloud-developed hardware acceleration devices. They provide overall system control and I/O virtualization passthrough. They run independently of the frontend host system and are powered independently.

  QingTian Cards enable secure boot and trusted measurements for the overall system based on a hardware root of trust. They also provide anti-tampering protection for firmware and accelerate hardware-based I/O encryption and decryption.
- QingTian Hypervisor is a lightweight hypervisor designed to provide powerful resource isolation, strong security, near-bare-metal high performance.

Dedicated QingTian Cards connect to the host system CPU through the standard PCI-Express interface and simulate various local and network resources as local resources of the host CPU through drivers. This prevents customers from configuring complex functions and ensures secure isolation between cloud infrastructure and customer applications. QingTian Cards also use dedicated ASIC hardware to process storage and network virtualization, which not only improves performance but also reduces costs. In addition, the QingTian architecture can greatly reduce the adaptation workload between the cloud infrastructure foundation and different computes, improving the iteration speed of new functions.

1.3 QingTian Threat Assumptions and Security Methods

Threat Assumptions

To provide tenants with full-stack sensitive workload isolation and data protection capabilities, the QingTian system assumes the following three threat modes in its security design:

- **Threat type 1: An adversary attacks tenant VMs by controlling the VMM.**

  Typical attack patterns include:

  - A malicious tenant on the cloud purchases an ECS and exploits a VMM zero-day vulnerability or side channel attack to perform a VM escape attack. As a result, the VMM is controlled and other tenant instances running on the same hardware are attacked.
  - Cloud service provider (CSP) internal personnel use valid credentials to remotely access hosts for deployment, change, commissioning, and diagnosis, and use attacker tools to read or tamper with sensitive data of tenant VMs.
- **Threat type 2: An adversary enters the data center to perform local physical attacks.**

  Internal personnel of the data center need to access physical devices in the data center due to hardware deployment, maintenance, and repair. Typical attack patterns include:

  - After stealing a hard disk, an adversary uses tools to access the hard disk data offline.
  - An adversary eavesdrops on all traffic data transmitted between physical network devices.
  - An adversary pre-installs, changes, or injects malicious firmware into the server mainboard system.
- **Threat type 3: An adversary attacks sensitive applications of tenants by controlling the guest OS.**

  Typical attack patterns include:
  - Attackers implant untrusted code through the software supply chain, exploit guest OS zero-day vulnerabilities or incorrect configurations to perform privilege escalation attacks, and use attacker tools to read or tamper with sensitive applications and their data after obtaining the root permission of the guest OS.
  - Internal personnel of customers use valid credentials to remotely access VMs for deployment, change, commissioning, and diagnosis, and use attacker tools to read or tamper with sensitive applications and their data running on VMs.

Security Methods

The QingTian system adopts the following principles and methods in security design to address the preceding three types of threats.

- **Defend against malicious use of VMM**
  - The QingTian system uses the frontend and backend separated VMM architecture to offload VM management and I/O virtualization to the backend QingTian Cards. This isolates cloud system management from tenant workloads.
  - Based on the minimum trusted computing base (TCB) design principle, the frontend QingTian Hypervisor only retains the basic code for virtualization running, greatly reducing the risk of VM escape.
  - Based on the unidirectional control flow of "ECS Control Plane -> QingTian Cards -> QingTian Hypervisor", only unidirectional connection initialization is allowed. The threat radius of escape attacks is limited level by level to enhance in-depth security defense.
  - Based on the mandatory secure boot and trusted measurement methods, integrity protection and exception detection are provided for the frontend and backend system firmware, boot system, and hypervisor.
  - Based on zero-privilege O&M, O&M APIs are provided to replace traditional SSH remote login to access servers. The OS has been streamlined, with the protocol stack, file system, network packet capture tool, and memory export tool all removed.
- **Defense against local physical attacks**
  - Based on data encryption, block storage encryption and Virtual Private Cloud (VPC) traffic encryption are supported to encrypt the I/O data related to tenant VMs after the data leaves QingTian compute nodes.
  - Based on the hardware-protected key method, data keys are distributed end-to-end from the KMS hardware to QingTian Cards hardware securely.

  - Based on hardware identity authentication and trusted measurement, untrusted hardware devices are prevented from being connected or mounted, and tampered system firmware is prevented from being booted.

  In addition, memory encryption and bus data encryption will be supported in next-generation servers to further improve the system security baseline.

- **Defense against malicious use of guest OSs**
  - Based on trusted computing, UEFI secure boot and QingTian Trusted Platform Module (TPM) are available on tenant VM instances to support standard trusted measurement and remote attestation methods, as well as integrity monitoring of guest OSs.
  - Based on the design method of isolating tenant VMs from the cloud system, the QingTian Enclave feature is provided for VM instances to isolate sensitive workloads in VMs from guest OSs.

    Sensitive workloads of tenants only run in the QingTian Enclave environment. Even if attackers completely control the guest OS, the confidentiality and integrity of the Enclave runtime environment are not affected.

1.4 QingTian System Components

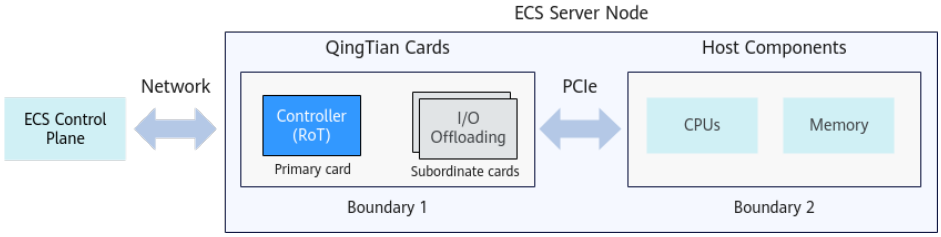
1.4.1 Overview

A QingTian ECS server consists of a host system and QingTian Cards.

- QingTian Cards are dedicated hardware components that run independently from the host system.
- The host system consists of the host CPU and memory, and runs customer's ECS VM instances or bare metal instances.

QingTian Cards and the host system are connected through the PCIe bus. They belong to two completely isolated system security domains.

Figure 1-2 QingTian system components



1.4.2 QingTian Cards

Logically, QingTian Cards consist of one primary card and multiple subordinate cards.

- Primary card: also called QingTian Controller. It manages all other components and firmware of the server system.

- Subordinate card: also called the I/O offloading card. It provides dedicated network acceleration, storage acceleration, management plane acceleration, and encryption offloading.

QingTian Cards contain all control interfaces required by the ECS service to provision and manage CPUs, memory, and storage for hosts. For the ECS service control plane, the central node only needs to connect to QingTian Cards, not to servers. Servers do not contain storage and network. All VM lifecycle management commands from the ECS service control plane are directly delivered to QingTian Cards. QingTian Cards operate front-end servers unidirectionally, such as creating or stopping VMs, hot plugging devices, and live migrating VMs.

QingTian Cards also provide all I/O interfaces for servers to interact with external systems, including VPC network interfaces and EVS block storage interfaces. For servers, all components (logical inbound or outbound) that interact with the external world are implemented through QingTian Cards. QingTian Cards are connected to servers via PCIe and can be powered independently. The cards encapsulate Huawei-developed SPU chips for firmware startup and simplified OS running. QingTian Cards support hot upgrade of the OS and key virtualization components on the cads. The upgrade is independently from the firmware and system components on the host server and does not affect customer services and security protection functions.

QingTian Controller

The QingTian system supports secure boot based on the UEFI Secure Boot standard. After the server is powered on, QingTian Controller performs secure boot first. At this moment, the host system is waiting. The secure boot process of the system on a chip (SoC) in QingTian Controller is as follows:

1. Start boot ROM.
2. Verify the signature integrity of the firmware in the initial boot phase that is stored in the flash memory connected to QingTian Controller to complete the secure boot of QingTian Controller.
3. Verify the signature integrity of the QingTian Hypervisor image in the connected flash memory to extend the trust chain to the front-end host system.
   - If the image signature verification fails, an abnormal startup event is reported and the startup stops.
   - If the image signature verification is successful, the host system is notified to continue the secure boot.

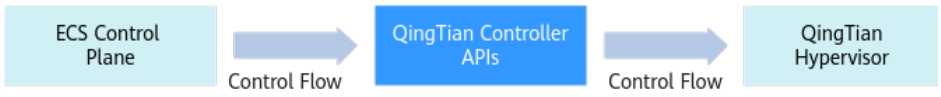
If the secure boot of the host system fails, an abnormal startup event is reported. If the startup is abnormal, the node is removed from the service node and does not run customer workloads.

QingTian Controller is also a security gateway that isolates the physical server and cloud service control planes. Cloud service control planes (including ECS/BMS, EVS, and VPC) are logically independent and use the microservice architecture. QingTian Controller abstracts the cloud service control planes as ECS Control Plane for interaction. The interaction follows a unidirectional control flow: ECS Control Plane -> QingTian Controller -> QingTian Hypervisor. Only unidirectional connection initialization is allowed. Any reverse initialization is considered abnormal.

As the only channel, QingTian Controller isolates the physical server from external control planes. All inbound and outbound traffic must be forwarded by QingTian Controller.

- QingTian Controller provides an mTLS-based bidirectional authentication communication link for the ECS service control plane to ensure end-to-end encryption of data transmission links.
- QingTian Controller also provides condition-based access control based on API context attributes to restrict each control plane component to only call the minimum set of APIs required for its services. In addition, the system records all API operation logs (including the source network context, identity context, call parameters, and timestamps) and supports real-time detection of abnormal API calls.
- QingTian Controller communicates with the ECS service control plane through a dedicated network. The inbound and outbound traffic on the control plane is completely isolated from the tenant traffic (such as EVS storage data traffic and VPC network traffic).

Figure 1-3 QingTian Controller interaction



I/O Offloading

The QingTian system has dedicated I/O offloading acceleration hardware. The hardware uses the same SoC and basic firmware architecture as QingTian Controller. It can accelerate offloading for networking and storage hardware, such as VPC and EVS block storage. The offloading acceleration hardware implements data encryption and acceleration for networking and storage using hardware encryption offloading engines and secure key storage integrated in the SoC.

NOTE

Huawei Cloud-developed VPC encryption: Standard security protocols IPsec and TLS are not applicable to communication encryption in large-scale, high-performance cloud data centers. Huawei Cloud has launched the cloud network CAE cryptographic algorithm based on its service security requirements to meet the encryption transmission requirements in multiple scenarios of Huawei Cloud networks, such as encrypted transmission between VMs in a given VPC and cross-site encrypted transmission in distributed clouds. Huawei Cloud supports secure, encrypted connections between all ECSs. For specific ECSs, dedicated offloading cards for VPC can be used to encrypt in-transit traffic between instances. By default, the CAE protocol uses the AES-256-GCM algorithm to automatically and transparently encrypt the in-transit traffic between instances. The encryption protocol supports anonymity, anti-replay, forward-secrecy, and post-quantum security.

Dedicated I/O offloading cards provide hardware acceleration for data key import and encryption/decryption algorithms based on end-to-end encryption. Standard cryptographic algorithms such as AES and encryption modes such as GCM/XTS are supported. The encryption keys used for EVS and VPC networks are present only in plaintext in the hardware key-protected subsystem of QingTian Cards. Huawei Cloud O&M personnel and any customer code running on the host system cannot access them. To avoid individual security issues in the key distribution system, the control plane system uses multiple key management components to distribute

multiple key materials independently and securely. The data plane does not need to perform key negotiation. Instead, at runtime it derives data keys from the multiple key materials delivered by the control plane and supports automatic key rotation on an hourly basis. This key distribution mechanism is more suitable for the cloud computing SDN architecture. It reduces performance overhead of key negotiation and enables communication encryption at greater scale and scope.

When the system runs in QingTian Hypervisor, the I/O devices provided by QingTian Cards are divided into multiple virtual functions (VFs) using the single-root I/O virtualization (SR-IOV) technology, and I/O devices can be directly connected to VMs. These VFs can be directly allocated to VMs so that the VMs can directly access hardware interfaces (such as network interfaces and storage controllers). In the transmission path, customer service data (transmitted for processing, storage, and hosting) is directly transmitted between ECSs and the virtual I/O devices provided by QingTian Cards, bypassing the hypervisor layer to achieve hardware-level data passthrough. Based on to the principle of minimizing the attack surface, this solution ensures that the I/O path only involves VMs, VF hardware, and physical devices. By minimizing the dependency on software and hardware in the I/O path, it delivers higher security and near-bare-metal performance.

1.4.3 QingTian Hypervisor

QingTian Hypervisor provides ultimate isolation and security for tenant ECSs through lightweight design, minimum attack surface, anti-tamper design, and hot upgrade.

Lightweight Design

Unlike traditional Type-1 Hypervisor, QingTian Hypervisor is designed to be lightweight.

- Full offloading architecture: The traditional management plane and I/O data plane are offloaded to QingTian Cards. QingTian Hypervisor only retains basic virtualization and device passthrough capabilities, and provision 100% server resources to tenant VMs. The management plane of QingTian Cards manages the lifecycle of VMs running on QingTian Hypervisor through the vsock secure channel. All virtual disks and virtual network interfaces are presented as standard virtio-PCI devices through QingTian Cards. The devices ensure performance and support flexible hot swap and live migration.
- Huawei Cloud EulerOS 2.0: An in-house lightweight, stateless virtualization OS. It removes all kernel modules and software packages that are irrelevant to virtualization, and only retains the components and modules necessary for running hypervisor. The system is compact, easy to transfer, and supports quick fixing of kernel vulnerabilities.
- VRAM: an in-house pageless lightweight memory management system that discards the traditional memory paging management. It reduces the management overheads by dozens of times while maintaining the memory compatibility of VMs.

Minimum Attack Surface

Compared with the traditional hypervisor, the lightweight QingTian Hypervisor also considers the impact of various external attack sources on the virtualization

data plane. QingTian Hypervisor minimizes the attack surface using the following technologies:

- Minimal TCB: The software code has been streamlined, only retaining the basic code for virtualization.
- No network: Network functions are removed. QingTian Hypervisor interacts with QingTian Cards only through the vsock secure channel to further reduce the attack risks on the management plane.
- No storage: The server does not have local disks, disk file systems, or configuration files. Logs and monitoring data are periodically recorded to the cloud through APIs. No status data can be edited or modified externally.
- CPU pinning: VMs are bound to dedicated CPUs, eliminating the need for QingTian Hypervisor to schedule CPUs. This avoids overhead of context switching and mitigates side-channel attacks.
- Strong isolation: Hardware-assisted virtualization and Huawei-developed VRAM memory management ensure that VMs cannot access each other's memory and I/O resources.

Anti-tamper Design

With the streamlined, lightweight software package, QingTian Hypervisor performs trusted verification during the secure boot of the host. QingTian Controller uses the authenticated encryption with associated data (AEAD) algorithm to protect the confidentiality and integrity of sensitive configuration data during the boot process. The configuration data is decrypted only when the encryption and decryption context is in the expected trusted environment. At the runtime of QingTian Hypervisor, the memory file system is configured to be read-only, and trusted audit is enabled to prevent VM escape or tampering by external software. The QingTian Hypervisor software package upgrade also needs to be verified through CRC and certificates to ensure that the software package is not tampered with during transmission.

System Hot Upgrade

Traditional hypervisor system software upgrades require shutdown or service migration before deployment. This may interrupt customer services and result in low upgrade efficiency. QingTian Hypervisor is updated periodically. To meet upgrade requirements in different scenarios, QingTian Hypervisor provides comprehensive secure hot upgrade, including function-level hot patches, component-level hot replacement, and in-place hot upgrade of the entire hypervisor system. Customer services are almost unaware of the upgrade. Thanks to the key capability of in-place hot upgrade of QingTian Hypervisor, large-scale parallel upgrades can be quickly performed within a cluster without service migrations and host reboot. Software versions can be released quickly. Throughout the upgrade process, the system continues to enforce complete security policies and defense capabilities.

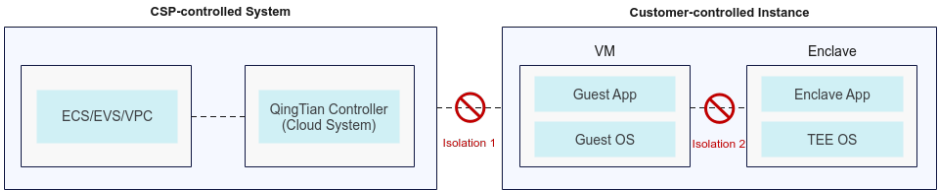
1.5 QingTian Confidential Computing

1.5.1 Overview

QingTian confidential computing aims to use dedicated hardware and related firmware as the root of trust to protect customer application code and data being processed from external access.

- Confidentiality protection: Ensures that customer data and code are not accessed by CSPs' internal personnel, cloud systems, and customers' internal personnel or VM administrators.
- Integrity protection: Ensures that customer data and code are not tampered with by CSPs' internal personnel, cloud systems, and customers' internal personnel or VM administrators.

Figure 1-4 QingTian confidential computing



To achieve these objectives, QingTian confidential computing provides security isolation in two dimensions:

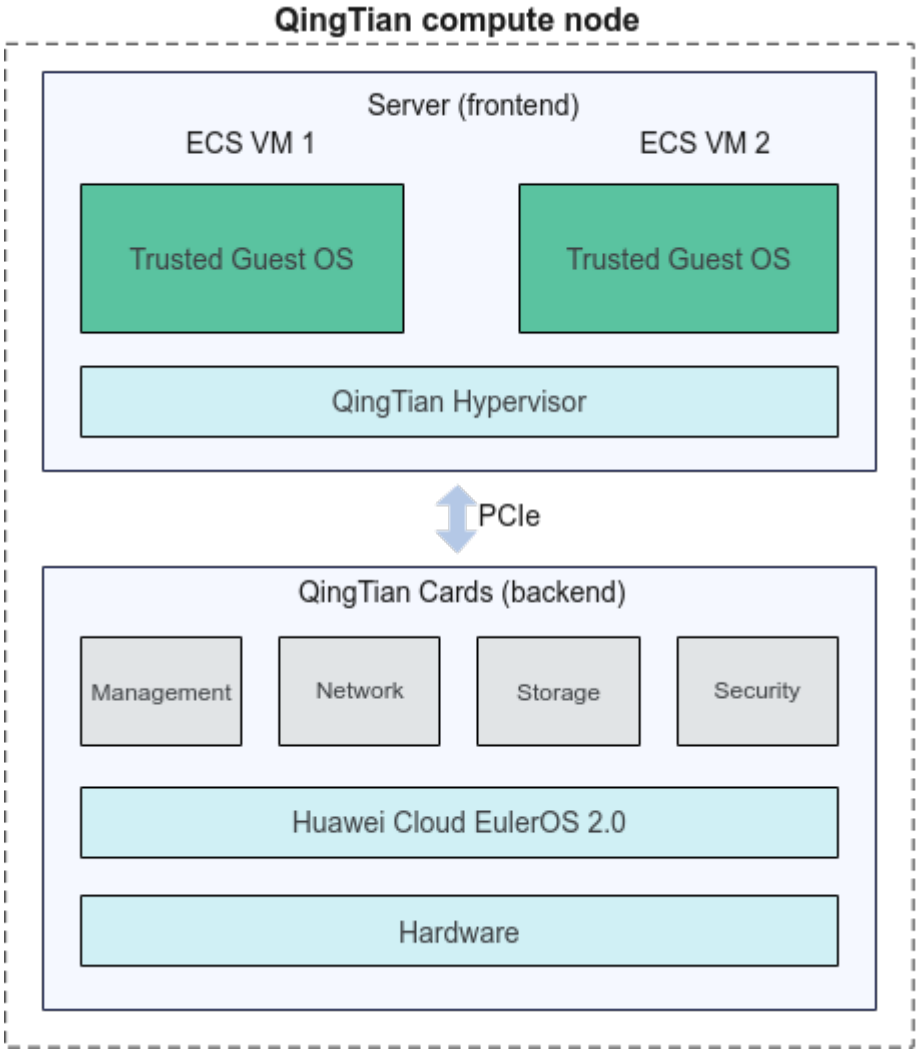
- Isolation dimension 1: Isolate customer data and code from CSPs' internal personnel and cloud system software.
- Isolation dimension 2: Isolate customer data and code from customers' internal personnel and untrusted guest OSs.

1.5.2 Isolation Dimension 1

One of the initial design objectives of the QingTian system is to support secure isolation between customer workloads and the CSP cloud infrastructure, including QingTian bare metal instances and QingTian VM instances.

- For QingTian bare metal instances, QingTian Cards are completely isolated from the host system. Because the host system does not run QingTian Hypervisor, customers exclusively access the underlying mainboard system and use related CPU hardware features (such as Arm TrustZone).
- For QingTian VM instances, QingTian Hypervisor provides strong isolation similar to that of bare metal instances through lightweight design, minimum attack surface, anti-tamper design, and hot upgrade.

Figure 1-5 Isolation dimension 1



The QingTian system uses the following methods for isolation in this dimension to enhance security of QingTian VM instances:

- Strong isolation: The QingTian system uses a frontend and backend separated VMM architecture, where the frontend and backend are physically isolated based on the PCIe bus. The frontend hypervisor isolates memory and I/O access between VMs based on hardware-assisted virtualization and in-house VRAM memory management. QingTian Cards support VM passthrough access to hardware devices through SR-IOV. In addition, CPU pinning enables the binding of VMs to dedicated CPUs, eliminating the need for QingTian Hypervisor to schedule CPUs. This avoids overhead of context switching and mitigates side-channel attacks.
- Escape prevention: Based on the minimum TCB design principle, QingTian Hypervisor code has been streamlined, retaining only the basic code required for virtualization. There are no network protocol stacks, local disks, configuration files, or SSH management tools. QingTian Hypervisor has a code volume less than 1% of traditional virtualization management systems, significantly lowering the VM escape risk.

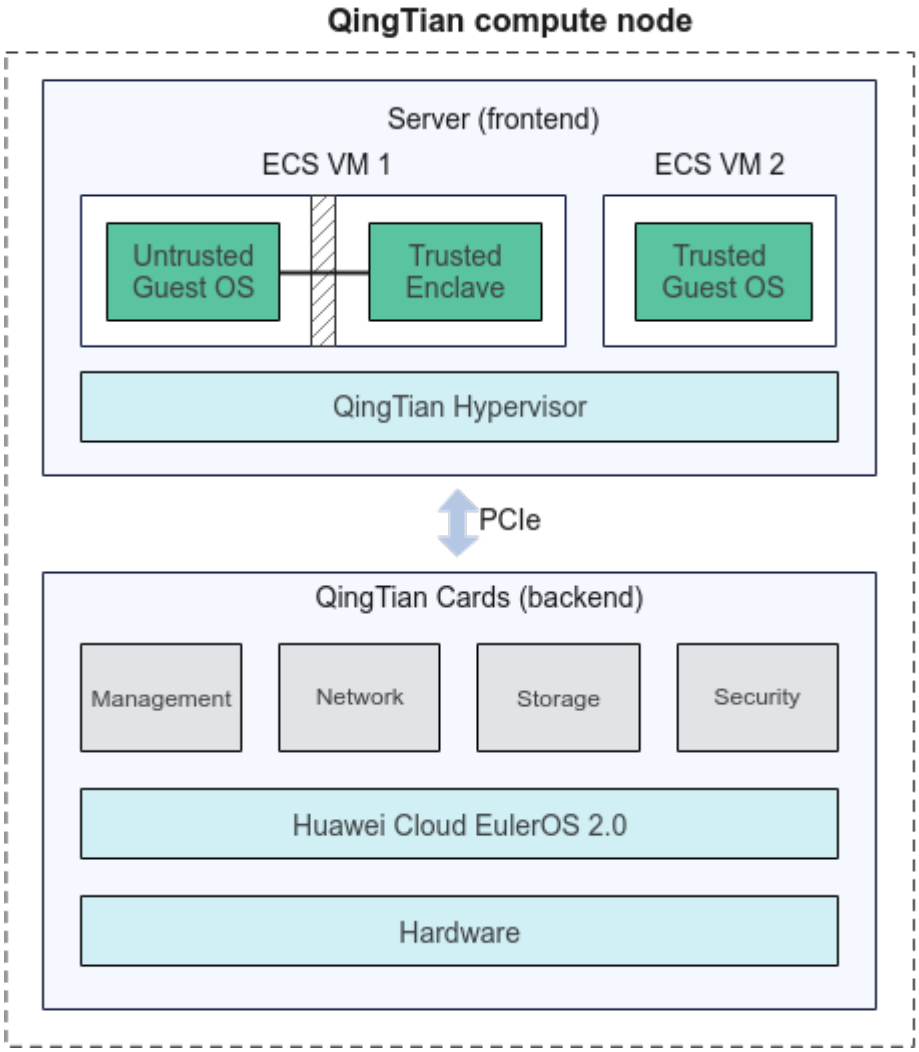
- Anti-tampering: The QingTian system uses forcible secure boot and trusted measurement. QingTian Controller performs secure boot first to ensure that the boot environment meets the expectation. Then, it verifies the integrity of the QingTian Hypervisor image file, and boots the host system to start QingTian Hypervisor.
- Key protection: QingTian Cards have an independent hardware security module. It uses the hardware-protected identity authentication to establish trusted access with the ECS control plane. This can prevent node identity spoofing caused by software credential leaks. QingTian Controller uses the hardware security module to protect the key materials required for volume encryption and VPC encryption, derives data keys in the hardware environment, keeps data keys within the hardware.
- Zero-privilege access: QingTian Hypervisor does not provide any remote login methods. Cloud service O&M personnel can only use O&M APIs for remote diagnosis. No internal personnel can obtain system privileges to access the memory data of customer instances.

1.5.3 Isolation Dimension 2

On the cloud infrastructure platform, providing integrity protection for the guest OS of VM instances has become a new security baseline. Currently, ECS provides UEFI secure boot and QingTian TPM features to support trusted boot and remote attestation that meet TCG standards. QingTian TPM, provided by the QingTian system, is a virtual device that complies with the TPM 2.0 specifications. The OS generally uses TPM to provide security features like disk encryption (such as Windows BitLocker) and data anti-tampering (such as Linux DM-Verity).

However, general-purpose guest VM OSs (such as Rich OS) usually have a large TCB, which often leads to a large attack surface. In addition, there are still many challenges in protecting the integrity of the guest OS at runtime. In this isolation dimension, we refer to the design method of isolating tenant VMs from the cloud system in isolation dimension 1 to use QingTian Enclave to run sensitive applications of tenants. It isolates the guest OS from the trust boundary of QingTian Enclave and completely isolates the runtime environment of sensitive applications from that of the guest OS. This design ensures that security threats in the guest OS do not affect the security of applications and data in the Enclave environment.

Figure 1-6 Isolation dimension 2



QingTian Enclave is an isolated VM runtime environment created from an ECS. It connects to the parent instance through a dedicated vsock secure channel. Based on the assumption that the guest OS of the parent instance is untrusted, QingTian Enclave uses the following methods to enhance security in this isolation dimension:

- Strong isolation: QingTian Enclave and the customer's parent instance are isolated based on hardware-assisted virtualization and in-house VRAM memory management. The Enclave and parent instance do not share physical memory or CPU cores. They are connected only through a hypervisor-protected dedicated local channel vsock. Even if the guest OS of the primary instance has security vulnerabilities or the super administrator is attacked, the attacker who controls the guest OS of the primary instance cannot access the code and data in the QingTian Enclave environment.
- Minimum attack surface: QingTian Enclave does not support network interface attachments and SSH interactive access, and does not provide network interfaces and persistent storage. A QingTian Enclave OS is a Huawei Cloud-developed security OS that has been streamlined to the minimum requirements. Customers can also customize their own Enclave OSs.

- Anti-tampering: When QingTian Enclave is started, QingTian Hypervisor verifies the digital signature of the Enclave image and measures the Enclave image file and the public key certificate of the digital signature. The measurement results are saved to QingTian Security Module (QTSM). QTSM provides TPM-like trusted measurement and remote attestation. The difference is that QTSM redefines the trusted measurement attributes and attestation security protocols based on ECS scenarios.
- Key protection: QTSM runs in the isolated compute environment provided by QingTian Cards, generates a random attestation key pair based on the TRNG hardware engine, and applies for an attestation public key certificate from QingTian Attestation PKI. After the hardware-enhanced identity authentication is successful, QingTian Attestation PKI issues an attestation public key certificate to QTSM. QTSM also supports hourly rotation of attestation certificates, further reducing the risk of key leakage.

QingTian Enclave supports remote attestation protocols. When establishing trust with external parties, Enclave applications can provide cryptographic attestation of Enclave identity and runtime environment measurements to the parties through the attestation protocol. Huawei Cloud Key Management Service (KMS) and Identity and Access Management (IAM) inherently support QingTian Enclave attestation. QingTian Enclave application developers can use the open-source Enclave SDK to access KMS APIs. These APIs allow them to obtain data encryption/decryption keys or secure random numbers and ensures E2E security. IAM administrators can use preset IAM authorization policies or guardrail policies to enforce attestation-based conditional access control on KMS APIs.

In addition, QingTian Enclave is a developer-friendly platform in terms of usability and application compatibility. Developers can easily develop QingTian Enclave applications without CPU microarchitecture expertise and advanced cryptography knowledge. QingTian Enclave supports both x86 and Arm architectures. Developers can use their familiar language frameworks to build QingTian Enclave images using container images.

QingTian Enclave enables customers to create a highly isolated and enhanced compute environment within the ECS VM environment, so that customers can isolate their system components based on different trust levels. QingTian Enclave has been favored by many cloud customers since its launch. Production applications built based on QingTian Enclave include vHSM, vault credential management, and MPC wallets. For the cloud native confidential container solution, we also support configuring the QingTian Enclave device plugin in Kubernetes so that customer pods and containers can access the QingTian Enclave device driver. The device plugin applies to Cloud Container Engine (CCE) or customer-managed Kubernetes nodes. In addition, we provide a variety of QingTian Enclave open-source tools (such as qproxy) and security solutions to help more customers smoothly migrate to the QingTian Enclave environment without reconstructing application code and building systems.

NOTE

Application scenarios of QingTian Enclave: QingTian Enclave provides an extremely isolated runtime environment for applications with the minimum attack surface. This environment does not allow elastic network interface (ENI) attachments and storage volume mounting, neither support network protocol stacks and persistent storage. It can access external networks only through the vsock channel connected to the primary instance and the network proxy of the primary instance. Even if the guest OS of the primary instance is completely attacked, the application code and data security in the Enclave environment are not affected. If users want to attach ENIs or mount storage volumes in the isolation environment, or intend to access GPU devices, QingTian Enclave is not a good choice. In this case, ECSs that support QingTian TPM are recommended.

1.5.4 Cryptographic Attestation

QingTian Enclave Attestation

QingTian Enclave supports boot measurement and remote attestation. The measurements of QingTian Enclave is a group of hash values obtained by using the standard trusted measurement operations (ExtendPCR). The hash values are stored in the platform configuration registers (PCRs) of QTSM. QTSM supports a maximum of 32 PCR measurement attributes. PCR16 to PCR31 can be customized by Enclave applications. PCR0 to PCR15 are reserved by the QingTian system. The currently supported measurement attributes include PCR0 (QingTian Enclave image file), PCR3 (IAM Agency URN), PCR4 (ECS instance ID), and PCR8 (signing certificate of the QingTian Enclave image file).

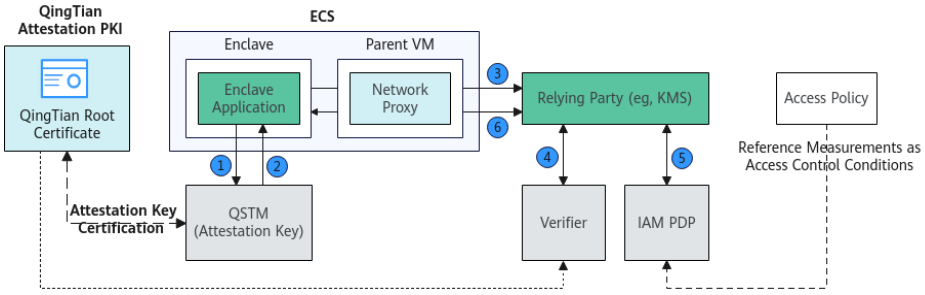
Enclave applications can obtain an attestation document of the current Enclave environment through QTSM. An attestation document includes the PCR measurement attribute list, QingTian PKI certificate chain, cryptographic algorithm declaration, and user-defined parameters. QingTian Enclave attestation documents support various user-defined parameters, including pubkey (user-defined public key signed by QTSM), nonce (one-off data to prevent replay attacks), and user_data (any user-defined data). It can support multiple user-defined security protocols (such as secure key negotiation and end-to-end encryption) between Enclave applications and external entities to provide complex application security solutions.

Huawei Cloud KMS and IAM support the QingTian Enclave attestation protocol. Tenants' IAM administrators can set condition-based authorization policies to achieve specific security control objectives. For example, only specified Enclave applications can call specified KMS APIs (such as generating random numbers, encryption, and decryption). The figure below shows how an Enclave application uses an attestation document that contains a custom public key to access the KMS Decrypt API to decrypt ciphertext data.

1. The Enclave application randomly generates an RSA key pair (private key and public key) and uses the public key as a parameter to call the QTSM API to generate an attestation document.
2. QTSM calculates the attestation document (containing the public key) and returns it to the Enclave application.
3. The Enclave application carries the attestation document and ciphertext data blob and calls the KMS Decrypt API.

4. KMS uses Verifier SDK to verify the attestation document.
5. KMS provides the verification result to IAM Policy Decision Point (PDP) for access control decision-making.
6. KMS decrypts the blob after the access control policy check is passed, re-encrypts the decryption result using the public key carried in the attestation document, and returns the result. After receiving the response, the Enclave application uses the private key generated in **1** to obtain the plaintext.

Figure 1-7 Decrypting ciphertext data



QingTian TPM Attestation

When creating an ECS, you can configure the UEFI secure boot mode and enable QingTian TPM. QingTian TPM is a virtual device provided by the QingTian system for ECSs and complies with the TPM 2.0 technical specifications. QingTian TPM provides measurement boot and remote attestation. You can obtain the signed PCR values from QingTian TPM and use them to prove the integrity of ECSs to remote entities. QingTian TPM can also generate keys for encryption or signing. The keys generated by QingTian TPM can be used to provide device attestation to relying parties.

ECS Identity Attestation

- **Using an instance identity document**

  Each ECS you start has an instance identity document that provides metadata about the instance, including the instance specifications, instance ID, image ID, account ID, private IP address, and creation time. You can use the instance identity document to verify instance attributes. Applications running on ECSs can obtain the instance identity document and digital signature of the instance identity document through Instance Meta Data Service (IMDS). When an application needs to send the instance metadata document to a remote entity (relying party) for verification, the instance metadata document and its digital signature must be provided. When obtaining the digital signature of the instance metadata document from IMDS, you can provide a custom audience parameter (for example, a one-time challenge value) to prevent replay attacks.

  The instance metadata document and digital signature can be regarded as the default birth certificate provided by the ECS service for each ECS instance. When an application in an instance needs to interact with a remote entity, the birth certificate can be used for initial identity authentication of the instance. With the "birth certificate", based on the security principle of transitive trust, the application can use the initial identity to obtain application-related access

credentials. This method can effectively solve identity security issues (such as hardcoding static credentials in application configuration files).

- **Using an IAM agency for an instance**

  You can configure an IAM agency when creating an ECS. An IAM agency for an instance is a virtual identity created by an IAM administrator. It represents the IAM identity for ECSs to access cloud service resources. An IAM agency does not have static credentials. This effectively reduces the risk of static credential (or long-term credential) leakage. An application running on an ECS can obtain a temporary security token of the IAM agency for the instance via IMDS. The security token is issued by the Security Token Service (STS) and represents an identity session of the IAM agency for the instance. After the IAM agency for the instance is authorized by IAM, the application can use the temporary security token of the IAM agency for the instance to access authorized cloud service resources (for example, OBS objects).

  The IAM agency for the instance can be regarded as a custom machine identity provided by the tenant's IAM administrator for the ECS. When an application running on the ECS needs to access authorized Huawei Cloud service resources or APIs, the application can directly use the temporary security token of the machine identity without hardcoding static credentials (such as AK/SK) in the application code or configuration file, thereby preventing static credential leakage.

  NOTE

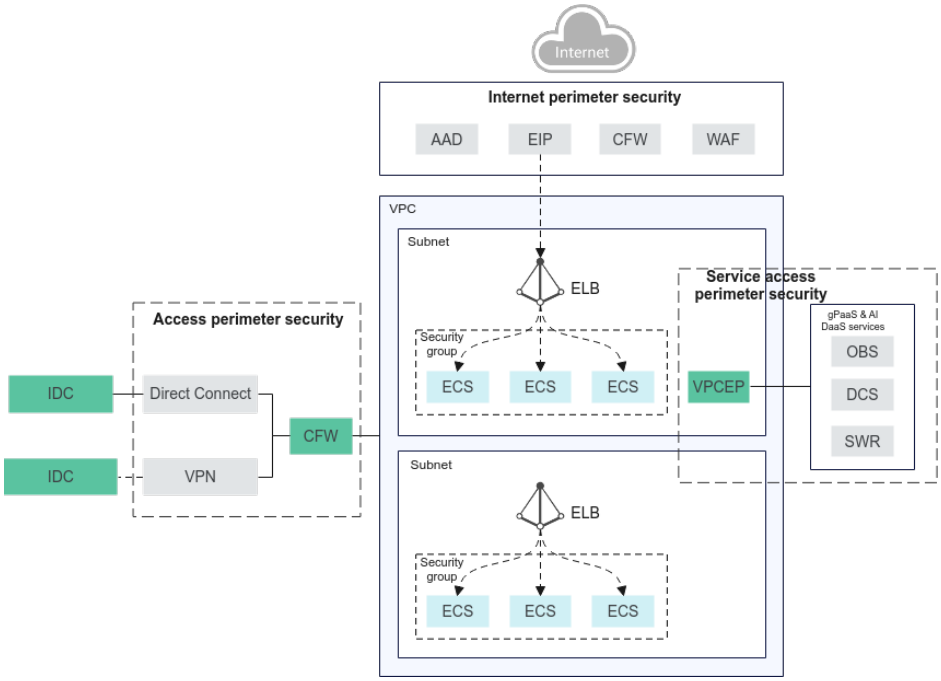
  ECS IMDSv2 is recommended for accessing IMDS. Compared with IMDSv1, IMDSv2 uses the PUT method and dynamic token to initialize IMDS sessions. This effectively reduces SSRF vulnerability risks of tenant web applications and potential security risks caused by incorrect configurations of WAF, reverse proxy, layer 3 firewall, or NAT. IMDSv2 can significantly enhance the in-depth security defense capability of tenant ECSs.

1.6 From Physical Isolation to Logical Isolation

Isolation is a cornerstone of cloud platform security. The Huawei Cloud QingTian system not only provides tenants with physical isolation of the runtime environment, but also extends logical isolation as part of its TCB, including VPC network isolation, multi-tenant resource isolation, and cross-cloud service access isolation.

VPC-based Network Isolation

Figure 1-8 VPC-based network isolation



VPC is a cornerstone of tenant security. It provides a completely isolated network space for customers to deploy services on the cloud. A VPC is a secure network environment by default. Customers can build necessary network channels as required while ensuring security. VPCs communicate with external networks using the following connection methods:

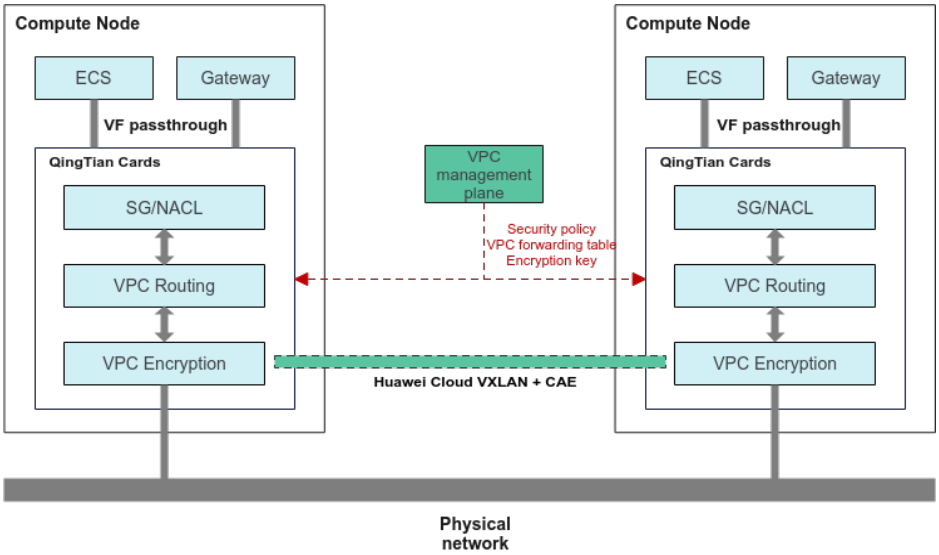
- Internet access: Security measures such as Anti-DDoS Service (AAD), Web Application Firewall (WAF), and Cloud Firewall (CFW) are used to ensure the security of Internet access through Elastic IP addresses (EIPs).
- Hybrid cloud access: CFW and other methods are used to safeguard on-premises equipment room access through Direct Connect and Virtual Private Network (VPN).
- Cloud service access: VPC Endpoint (VPCEP) is used to access cloud services, and VPCEP policies are used to ensure secure access to cloud services.

VPC provides two basic network security access control capabilities: security group and network access control list (ACL). Both are Layer 4 stateful network security capabilities.

- Security group: configures access policies for inbound and outbound traffic over ENIs or ENI sets.
- Network ACL: configures access policies in the inbound and outbound directions for subnets, including the source and destination IP addresses, ports, and protocols, and associates with specified subnets.

VPC also supports network encryption. For a VPC with encryption enabled, the traffic between compute resources and gateways within the VPC is transmitted using encrypted packets. This ensures that all customer traffic is encrypted and cannot be cracked or listened to by cloud providers.

Figure 1-9 VPC network encryption



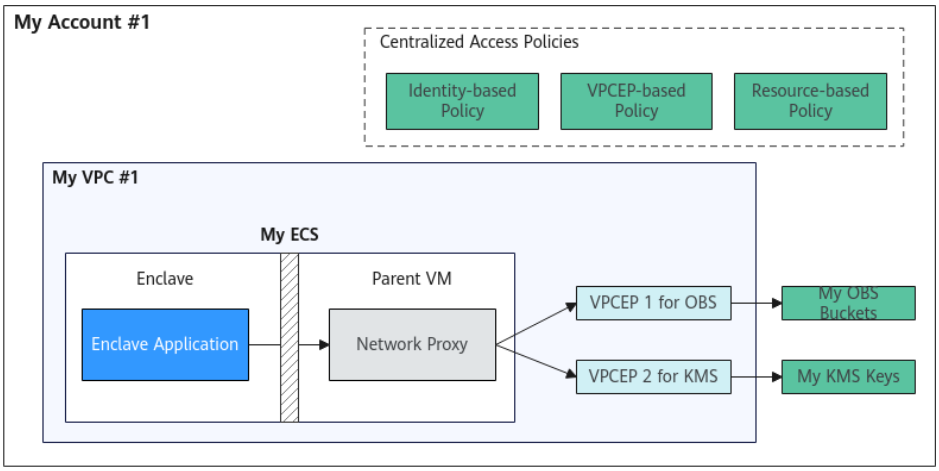
The QingTian hardware system enables VPC to provide network isolation, traffic encryption, and network access control based on security groups and network ACLs. The VPC management plane delivers the entries required by the network subsystem to QingTian Cards. QingTian hardware virtualizes ENIs through VF passthrough and connects to compute resources (such as ECSs, BMSs, and CCI pods) for intra-VPC forwarding.

- Security access control module: To control access through security groups and network ACLs, QingTian hardware identifies the ENI and subnet of each packet based on its associated VF, and then finds the corresponding security groups and network ACL security policies. After policy evaluation, QingTian hardware determines whether to forward the packet and generates session entries required by the stateful firewall to ensure that subsequent packets are correctly forwarded.
- VPC routing module: To enforce VPC-isolated packet forwarding, QingTian hardware searches for the routing entries dedicated to the customer's VPC based on the ENIs and subnets of the packets, encapsulates the packets in a VXLAN tunnel, and sends the packets. The destination QingTian hardware locates the VPC-specific routing tables of the packets based on the encapsulated information in the VXLAN and forwards the packets. Throughout the entire forwarding process, forwarding is isolated by searching for VPC-specific routing tables.
- VPC encryption module: To encrypt VPC traffic, QingTian hardware finds the encryption key dedicated to the VPC based on the packets' VPC attributes, and adds an encrypted packet header to the VXLAN tunnel to encrypt all packets, including the MAC address, IP address, and other forwarding information, as well as additional information about the VPC of the packets in the VXLAN tunnel. The encryption key is shared among compute nodes in a given VPC. The destination QingTian hardware can find the key based on the key ID, decrypt the key, and continue the subsequent forwarding process.

Multi-Tenant Resource Isolation for Cloud Service Access

In a tenant VPC, customer applications deployed on Enclaves (or VMs) may depend on API calls of multiple cloud services, such as OBS or KMS. However, these cloud services are deployed outside the tenant VPC. To allow customer applications to access cloud service APIs, Huawei Cloud uses VPC endpoints to break the VPC network isolation perimeter. VPC endpoints use one-way access and endpoint policies to control the attack surface on the access path.

Figure 1-10 Multi-tenant resource isolation



By default, cloud service APIs support multi-tenant resource access control and isolation based on the IAM access control system. Huawei Cloud IAM provides tenants with centralized identity and access control management as well as unified access control policy language, and offers consistent access control decisions for all cloud service APIs. IAM PDP supports multiple types of access control policies, including VPCEP policies, IAM identity policies, and cloud service resource policies. IAM access control policies support a variety of condition attributes. Currently, more than 50 global condition attributes and more than 500 cloud service condition attributes are supported. These condition attributes include identity attributes, session attributes, runtime environment attributes, network attributes, API context attributes, and target resource attributes. Tenants can flexibly combine condition attributes to customize IAM policies to meet security control or compliance requirements.

The QingTian system ensures correctness and integrity of the runtime environment attributes and network attributes required by IAM PDP. The QingTian system registers metadata with IAM through IAM Context Provider. The metadata includes the context verification public key and context assertion schema. QingTian injects context assertions with digital signatures into API access links. IAM PDP verifies the integrity of the context and executes the access control policy evaluation logic related to API requests based on the trusted context assertions.

Multi-Tenant Resource Isolation Across Cloud Services

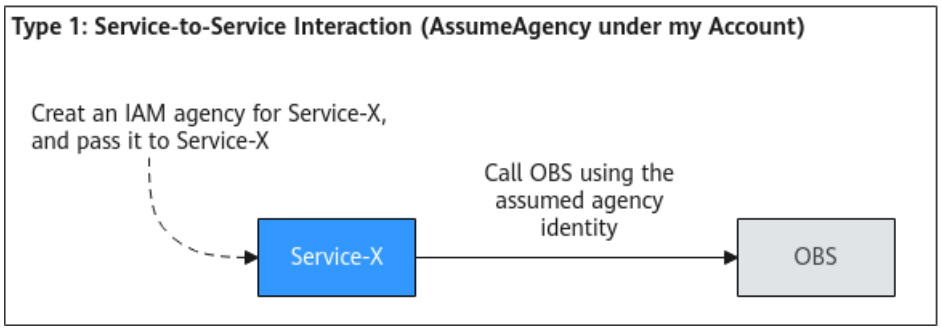
If your account subscribes to multiple cloud services, cross-service resource access may be involved. Cross-service resource access may cause the confused deputy problem. Inappropriate design of service access isolation may result in unauthorized access by other tenants.

Huawei Cloud solves the confused deputy problem at the protocol design level. By default, the subscribed service principal is restricted within the account domain of the tenant. Different cloud services in the same account domain are completely isolated. Access cross cloud services requires explicit authorization from the tenant.

Huawei Cloud IAM provides AssumeAgency and Impersonation security protocols to authorize access across cloud services in a given account domain.

- **Protocol 1: AssumeAgency**

Figure 1-11 AssumeAgency

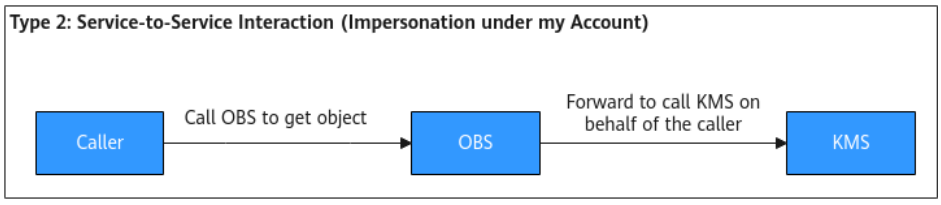


AssumeAgency is typically used when cloud service jobs need to be authorized to access downstream cloud service resources after users have gone offline. An administrator creates an IAM agency, specifies Service-X as the trusted principal in the trust policy, grants appropriate permissions to the agency, and passes the agency to Service-X.

By default, Service-X in the IAM agency trust policy indicates Service-X in your account domain (or organization domain). Service-X subscribed by other accounts is not trusted, and Service-X in other account domains does not have permission to assume the agency. By introducing the account domain (or organization domain), we can solve the confused deputy problem of cross-cloud service access at the protocol level and provide default multi-tenant security isolation.

- **Protocol 2: Impersonation**

Figure 1-12 Impersonation



Impersonation is typically used when the cloud service request processing logic needs to access downstream cloud service resources when users are online, such as the flow of "User -> OBS -> KMS". Administrators do not need to grant additional permissions to OBS. This protocol allows OBS to impersonate the identity and permissions of the requester to access the downstream cloud service KMS.

For example, a user attempts to call the OBS GetObject API to get the object data encrypted using the tenant KMS key. However, the OBS service principal was not authorized to access the tenant KMS key. In this case, OBS needs to impersonate the caller to call the KMS API.

Specifically, OBS provides API RequestProof of the caller to apply for a forwarding access token from STS, and then uses the token to construct a KMS API access request. API RequestProof contains the API name, API signature, and expiration time. It ensures security of impersonation. STS checks whether the API has an associated impersonation license and whether the API signature is valid, and then determines whether to issue the corresponding forwarding access token. The permissions of the forwarding access token obtained by OBS are the intersection of the permissions granted to the caller and the scope-down session policy associated with the API. The scope-down session policy associated with an API is designed based on the principle of least privilege. It can be registered in the production environment only after being reviewed by the cloud security team, IAM team, and cloud service team.

1.7 Zero-Privilege O&M

As mentioned in **QingTian Threat Assumptions and Security Methods**, cloud vendors' internal personnel may be potential attacks that pose threats to tenant data security through logical or physical access. The QingTian system introduces the zero-privilege O&M concept to eliminate such security risks by combining best technical practices and O&M security management.

- Zero SSH permission: O&M personnel cannot obtain control permissions on servers through common remote login methods. All O&M operations on servers are performed through APIs. All O&M APIs have strict identity authentication, authorization, recording, and auditing. These APIs do not allow O&M personnel to access customer data on servers.
- Zero packet capture tool: The OS of the server node has been streamlined, with common packet capture tools such as tcpdump removed, to prevent tenant I/O data from being listened to or stolen.
- Zero memory permission: VRAM, a QingTian memory management component, independently manages VM memory. The memory cannot be exported via traditional virsh dump.

These technical restrictions are built into the QingTian system. Even the system administrator with the highest permissions cannot bypass these controls and protections. As common login access is disabled, the production environment does not support in-place debugging. This is inconvenient for technical personnel, but we believe that this is a good trade-off for our customers. We must maintain high system quality and testing standards before production release.

Cloud infrastructure security also depends on the root key protection of multiple key systems. Root key protection needs to handle various threats, including potential threats from internal personnel with the highest permissions. Hardware security modules (HSMs) cannot avoid individual security issues. Only relying on this key protection technology cannot address increasingly severe attack challenges. The QingTian system uses the following methods and measures to protect the keys of key systems:

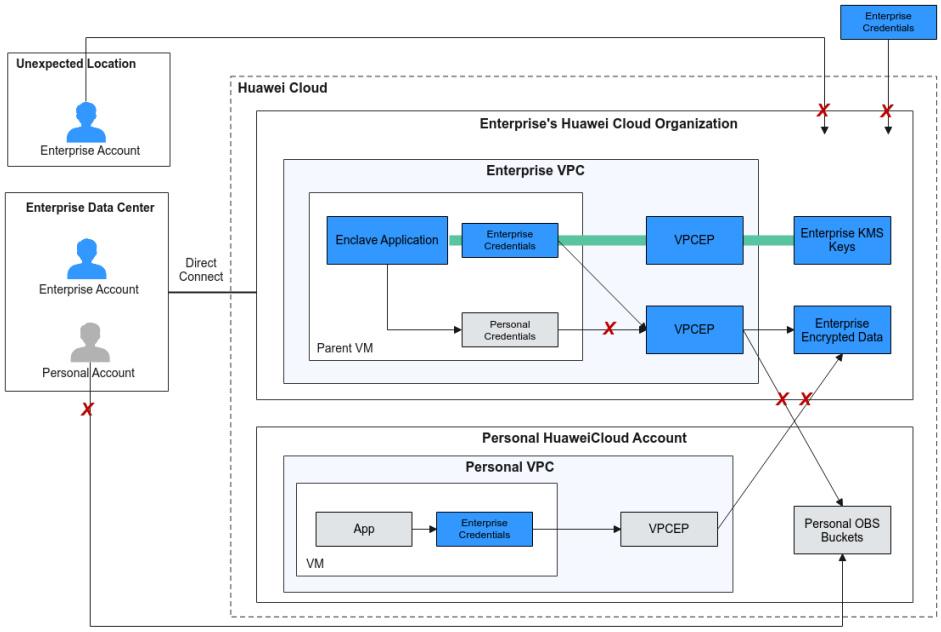
- Threshold digital signature: Some key systems use threshold digital signature algorithms to split signature keys and completely destroy the original keys after the splitting is complete. Once split, a number of *N* key fragments are separately held by multiple independent nodes. A number of *T* compute nodes can collaborate with each other to sign a message according to the threshold signature protocol. If the number of nodes is less than *T*, the message cannot be signed. Complete keys are not displayed throughout the signature process, and no single point holds the complete keys. This method not only improves the security of key management but also improves the availability of the system.
- Multiple key materials: Some key systems use multiple digital signature technologies (depending on multiple signature keys) for anti-tampering. Other systems synthesize data encryption and decryption keys based on multiple independent key materials. Each independent signature key or key material is protected by security hardware to prevent systematic security risks caused by the compromise of a single key.
- Key export prevention: Working keys are encrypted using hardware identity public keys and imported to the HSMs built into QingTian cards. They cannot be exported from the hardware in plaintext. Instead, they can only be used through module APIs with limited authentication and authorization.
- Fast key rotation: The system strictly ensures that the key rotation period is inversely proportional to the key usage frequency. Frequently used working keys (such as end entity certificate keys) are rotated on an hourly basis.
- Tenant-level key isolation: The system allocates completely independent random key materials to different tenants and derives tenant-level keys based on the key materials. Derived keys are only valid for single tenants, preventing global security impact caused by the compromise of a single key material.

Key protection has always been a challenge. We continuously focus on the latest cryptography technology progress (such as threshold cryptography technology), introduce these new technologies in cloud engineering best practices to improve security, and update the current engineering security best practices.

1.8 Case: Secure Cloud Migration of Financial Customer Data

Financial customers have strict requirements on data security. Based on the best practices accumulated by Huawei Cloud in the cloud security solution for financial customers, we propose a reference framework for financial customers to migrate data to the cloud.

Figure 1-13 Financial customer data security on the cloud



- **Objective 1: Core sensitive data only processed within the TEE confidential perimeter**

  All persistent data on the cloud is encrypted by default. For example, OBS bucket encryption, EVS volume encryption, and RDS database encryption are enforced through organization guardrails. Applications use an additional layer of encryption for core sensitive data. Data keys are protected by customer-controlled KMS master keys. IAM condition-based access control policies ensure that data keys can be decrypted and used only in the expected QingTian Enclave environment. Huawei Cloud KMS supports the QingTian Enclave Attestation protocol integration. It securely transmits data keys, random numbers, and decryption results from KMS to the Enclave environment, so that core sensitive data can be decrypted and processed only in the TEE environment.

- **Objective 2: Data boundary guardrails built in the production environment**

  Network boundaries are built based on VPC network configurations to prevent unexpected Internet inbound and outbound traffic, unexpected cross-VPC traffic, and unexpected cloud service traffic. In addition to network boundary control, the following identity and access control methods are used to build data boundary guardrails for the organization:

  – Identity-oriented data boundary guardrails: Use SCPs to define the permission boundaries of all IAM identities in an organization account to ensure that IAM identity credentials in the organization account can only use the VPC endpoints within the organization (public networks and other access paths are prohibited) and can only access cloud resources that belong to the organization account. After identity guardrails are built, even if IAM identity credentials are leaked, attackers cannot use the identity credentials to initiate access from the Internet or from other VPC endpoints. This design reduces the risk of credential leakage.

  - VPC endpoint-oriented data boundary guardrails: Use VPC endpoint policies to define permission boundaries to ensure that cloud service API requests passing through VPC endpoints only come from IAM identities in the organization account and can only access cloud resources that belong to the organization account.
  - Resource-oriented data boundary guardrails: Use resource authorization policies to define permission boundaries to ensure that all API requests for resources only come from IAM identities in the organization account and must pass through the organization's VPC endpoints (public network access paths are prohibited).
- **Objective 3: Secure identity access and credential leakage prevention**
  - A federated identity login solution is built based on the standard identity federation protocol. The solution prohibits employees from bypassing the local enterprise login system to use cloud accounts. To control tenant logins, the system restricts the network locations that allow for SSO logins and console access, and uses the local enterprise security gateway to enforce these restrictions. As a result, employees can access enterprise cloud accounts only from the enterprise intranet and cannot log in to personal accounts.
  - Application identity federation is built to allow tenant applications to use SAML and OIDC to securely exchange external tokens with Huawei Cloud IAM tokens. All IAM users are disabled. They are replaced by IAM agencies to eliminate the risk of long-term credentials leakage (such as AK/SK and login password). ECS IMDSv2 is forcibly used to access ECS instance identity signatures and IAM agency identity tokens. SCP policies are used to restrict the identity boundary guardrails of IAM tokens, eliminating risks caused by IAM token leakage outside VMs or VPCs.

1.9 Conclusion

This document provides an end-to-end security solution from the QingTian architecture system security to tenant security. This solution ensures that customers' sensitive services and data can run in a secure environment and prevents leakage of key sensitive data.

These security capabilities rely on the synergy of QingTian Cards, QingTian virtualization, and QingTian Controller, as well as the security of custom hardware.

Currently, all new ECS types are built based on the QingTian system, providing customers with all the security and benefits discussed in this document. Customers can select the most suitable instances to provide excellent security for their workloads based on the sensitivity.

1.10 Statements

This document describes the system security design of Huawei Cloud QingTian architecture.

- This document only provides security design information for reference.
- This document only applies to Huawei Cloud products and services. The content may change without notice.

The contents of this document do not constitute any warranty, statement, or commitment. Responsibilities and obligations of Huawei Cloud are stipulated in the legal agreements signed by customers and Huawei Cloud. This document does not constitute a legally binding agreement between Huawei Cloud and customers, nor does it constitute a change or modification to the agreement signed between Huawei Cloud and customers.