

Tencent Cloud CDN White Paper



Tencent
Cloud

[Copyright Notice]

© 2015-2020 Tencent Cloud. All rights reserved.

This document is copyrighted by Tencent Cloud and may not be reproduced, modified, plagiarized, or distributed, in whole or in part, without prior written consent of Tencent Cloud.

[Trademark Notice]



All trademarks related to Tencent Cloud services are the property of Tencent Cloud Computing (Beijing) Co., Ltd. or its affiliates.

All third-party trademarks are the property of their respective owners.

[Service Notice]

This document is intended to provide an overview of all or part of Tencent Cloud products and services, which are subject to change. The types and service standards of the Tencent Cloud products or services you purchased shall be governed by the commercial contract made and agreed upon between you and Tencent Cloud. Unless otherwise agreed by both parties, Tencent Cloud makes no guarantees or warranties, express or implied, about the content of this document.

Table of Contents

1	NETWORK STATUS AND CDN PRODUCTS	7
1.1	CDN PRODUCT OVERVIEW	7
1.1.1	<i>What Is CDN?</i>	7
1.1.2	<i>How CDN Works</i>	7
1.1.3	<i>General CDN Features</i>	8
1.2	TENCENT CLOUD CDN	9
1.2.1	<i>Tencent Cloud CDN Overview</i>	9
1.2.2	<i>Tencent Cloud CDN Deployment</i>	9
1.2.3	<i>Tencent Cloud CDN Advantages</i>	10
2	TENCENT CLOUD CDN SOLUTION ARCHITECTURE.....	12
2.1	SOLUTION OVERVIEW	12
2.2	CONTENT SYSTEM.....	12
2.2.1	<i>DiskTank3 Storage Engine</i>	12
2.2.2	<i>Two-Layer Cache Architecture</i>	14
2.2.3	<i>NWS Node Architecture</i>	15
2.2.4	<i>SCDN Platform</i>	15
2.3	MONITORING SYSTEM.....	16
2.3.1	<i>Multidimensional Monitoring System</i>	16
2.4	MANAGEMENT SYSTEM	17
2.4.1	<i>GSLB</i>	18
2.4.2	<i>Configuration Management System</i>	18
3	TENCENT CLOUD CDN PLATFORM SOLUTION.....	20
3.1	COMMON PLATFORM FEATURES.....	20
3.1.1	<i>HTTPS Optimization</i>	20
3.1.2	<i>Hotlink Protection</i>	24
3.2	WEBPAGE ACCELERATION PLATFORM.....	27
3.2.1	<i>Applicable Scenarios</i>	27

3.2.2	<i>Platform Architecture</i>	28
3.2.3	<i>Platform Features</i>	29
3.3	DOWNLOAD ACCELERATION PLATFORM	33
3.3.1	<i>Applicable Scenarios</i>	33
3.3.2	<i>Platform Architecture</i>	34
3.3.3	<i>Platform Features</i>	35
3.4	VOD ACCELERATION PLATFORM.....	38
3.4.1	<i>Applicable Scenarios</i>	39
3.4.2	<i>Platform Architecture</i>	39
3.4.3	<i>Platform Features</i>	40
3.5	LIVE VIDEO BROADCASTING ACCELERATION PLATFORM.....	45
3.5.1	<i>Applicable Scenarios</i>	45
3.5.2	<i>Platform Architecture</i>	46
3.5.3	<i>Platform Features</i>	47
4	CONCLUSION	50
5	RELEVANT LINKS	50

Table of Figures

Figure 1. Basic CDN workflow	7
Figure 2. Tencent Cloud milestones.....	9
Figure 3. Tencent Cloud global node distribution as of March 10, 2020.....	10
Figure 4. Tencent Cloud CDN platform architecture.....	12
Figure 5. DiskTank3 bypasses the file system	13
Figure 6. SCDN system architecture.....	16
Figure 7. Global distribution of Tencent Cloud CDN nodes as of 2020.....	17
Figure 8. GLSB architecture	18
Figure 9. SSL security test result of a website	21
Figure 10. Tencent Cloud SSL security test result	21
Figure 11. AES-NI effect comparison.....	23
Figure 12. False Start flowchart.....	23
Figure 13. URL hotlink protection flowchart	26
Figure 14. Webpage acceleration platform architecture.....	28
Figure 15. Effect of dynamic requests on loading speed	29
Figure 16. Static content origin-pull process	30
Figure 17. Dynamic acceleration request flowchart	31
Figure 18. Traditional TCP congestion control.....	32
Figure 19. Pornographic image detection system architecture	33
Figure 20. Download acceleration platform architecture.....	34
Figure 21. Dedicated DNS architecture	36
Figure 22. Dynamic content delivery recognition flowchart	38
Figure 23. VOD acceleration platform architecture.....	39
Figure 24. Steps for server-side upload.....	42
Figure 25. UGC video upload flowchart.....	44
Figure 26. Live video broadcasting acceleration platform flowchart	46
Figure 27. Standby mechanism.....	47

Figure 28. Multi-level elastic system.....	48
Figure 29. Live broadcasting key frame	49
Table 1. Applicable scenarios of the download acceleration platform.....	34
Table 2. Advantages of dedicated DNS	37
Table 3. Supported video formats	40
Table 4. Examples of preset container formats	41

1 Network Status and CDN Products

1.1 CDN Product Overview

1.1.1 What Is CDN?

Content Delivery Network (CDN) is a cost-effective, easy-to-use, and easy-to-implement solution to address issues such as content transfer, storage, and delivery over the internet. It is a layer of network architecture built on the internet, consisting of cache nodes distributed across regions to accelerate internet content delivery. These high-performance cache nodes store your content based on cache policies. When a content request is initiated, it will be routed to the closest node, effectively reducing access latency and improving availability.

1.1.2 How CDN Works

For example, if the domain name of your business origin server is "www.test.com" and it has been connected to CDN. When the user makes an HTTP request after the acceleration service starts, it will be processed as shown in Figure 1. Basic CDN workflow:

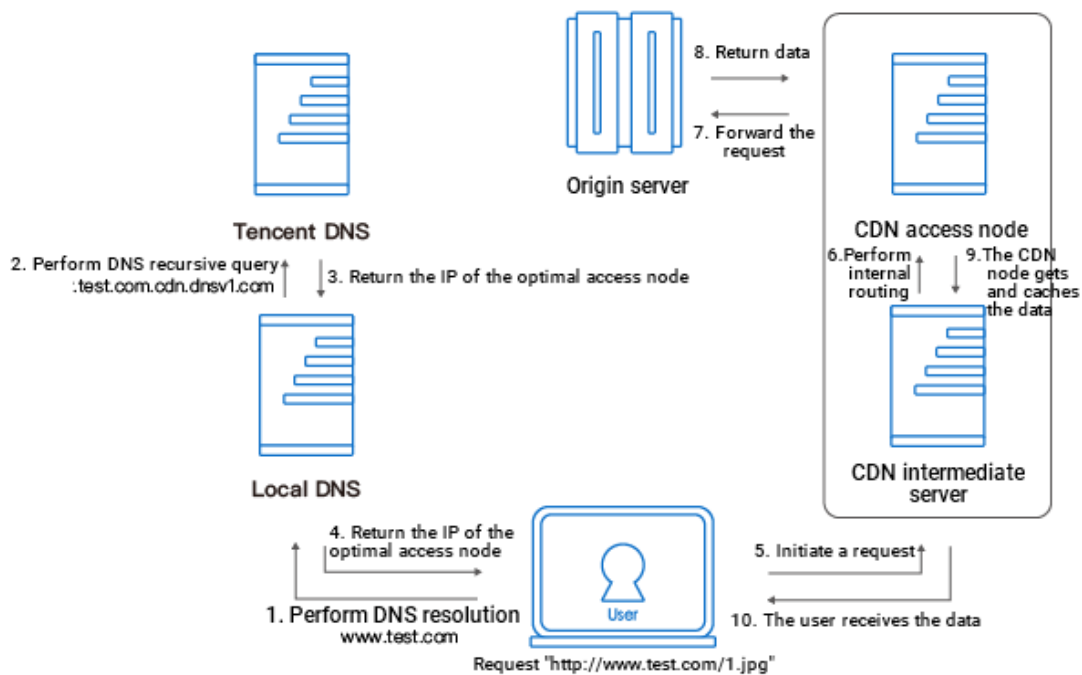


Figure 1. Basic CDN workflow

The process is detailed below:

- When the user initiates an access request for an image resource (e.g., "1.jpg") at "www.test.com", a domain name resolution request will be initiated to the local DNS;

- When the local DNS resolves "www.test.com", it will find that CNAME "www.test.com.cdn.dnsv1.com" has already been configured. The resolution request will be sent to Tencent DNS (GSLB), the scheduling system independently developed by Tencent Cloud, which will allocate the optimal node IP for the request;
- The local DNS receives the resolved IP returned by Tencent DNS;
- The user receives the resolved IP;
- The user makes an access request for "1.jpg" to the received IP;
- If the CDN node corresponding to the IP has cached "1.jpg", data will be directly returned to the user (10), and the process ends. Otherwise, the CDN node will initiate a request for "1.jpg" to the origin server (6, 7, and 8). After receiving the resource, the CDN node will cache it (9) based on the configured cache policy (please see "Setting Cache Validity Period" in User Guide) and return it to the user (10). The process ends.

1.1.3 General CDN Features

1.1.3.1 Improving Customer Experience

With the rapid development of the Internet, website content has evolved from text and graphics to the present form consisting of a large number of HD images and streaming media files. The overall resource consumption of the website increases, while requirements for network quality also rise. Moreover, Internet users not only demand more diverse website content, but also faster access. This brings an inevitable challenge to website operators.

1.1.3.2 Reducing Origin Server Costs

Currently, the video industry has entered stability with an environment of homogeneous competition. To boost customer retention, some websites purchased high-quality text or video content and cooperated with content providers to produce in-house content, which entail higher customer acquisition cost. Moreover, with increased network bandwidth and popularity of smart devices, users have higher requirements for display and video playback quality. If the website offers poor service quality, it may loss customers despite an initial investment. As website content keeps diversifying, the load on origin server also increases. Scaling up the origin server or implementing distributed deployment for user access often result in higher costs. In this case, CDN is widely recognized as the most cost-effective solution.

1.1.3.3 Protecting Network Security

Most websites have severe security risks, and Internet attacks have become increasingly common. For instance, DDoS and web attacks are very complicated and attackers often use multiple protocols, making network defense even more difficult.

In addition to attacks, content security is also a problem. To attract users, some websites and

companies purchase copyrights for movies and TV series. Since 2016, the value of intellectual property (IP) has been increasingly hyped. For websites that lack the financial support to purchase copyrights, they often plagiarize content through means such as hotlinking. These illegal activities remain in place despite repeated legal actions against copyright infringement. Tencent Cloud CDN aims to resolve this issue by collaborating with content providers.

1.2 Tencent Cloud CDN

1.2.1 Tencent Cloud CDN Overview

Based on the CDN platform independently developed by Tencent, these open-source CDN products offer a complete set of content delivery services for Internet businesses. Tencent Cloud aims to bring its diverse experiences and technical expertise in CDN acceleration to internet service providers.

Tencent Cloud started the development of CDN products in 2007. In more than a decade, it had gradually built a series of acceleration platforms for static content, game download, UGC, streaming media, dynamic content and more. These platforms serve a wide range of Tencent businesses, such as QQ, WeChat, Tencent Video, QQ Music, Tencent Games, mobile apps, Tencent Cloud, and QTalk. Since 2013, CDN has been opened to external users on Tencent Cloud. Backed by years of experiences and growth, global CDN products and solutions have thus been launched. In the future, Tencent Cloud CDN will open up more service capabilities to provide acceleration services that better meet the needs of internet businesses.

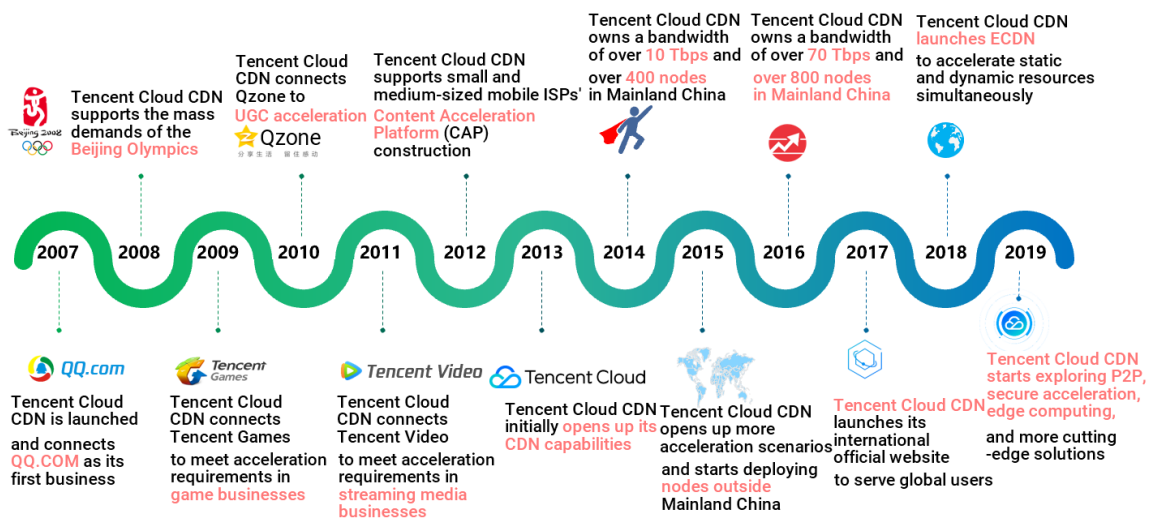


Figure 2. Tencent Cloud milestones

1.2.2 Tencent Cloud CDN Deployment

As of February 2020, Tencent Cloud CDN had deployed over 1,100 cache nodes across mainland China, and over 1,000 in more than 50 countries and regions worldwide. Backed by more than 120 Tbps of bandwidth, these cache nodes cover thousands of small- and medium-sized ISPs around the globe. Tencent Content Acceleration Platform (CAP), which is built in mainland China, connects the

network platforms of ISPs with Tencent businesses to provide high-quality Tencent content deployed on CAP to ISP users. These cache node resources guarantee stable and high-quality services.

CDN nodes are generally divided into two categories: cooperative proxy nodes and self-deployed network nodes. To guarantee the OPS quality of CDN network from the very first level (i.e., hardware), Tencent Cloud CDN has invested heavily in building high-availability and high-load data delivery nodes and data relay centers.



Figure 3. Tencent Cloud global node distribution as of March 10, 2020

1.2.3 Tencent Cloud CDN Advantages

- Stable, reliable and independently developed CDN nodes

Tencent Cloud CDN has over 2,100 global nodes offering a combined bandwidth of over 120 Tbps used by more than 600 million users every day.

It enjoys the same acceleration service as QQ, WeChat and Tencent Video.

- Optimal acceleration results

Tencent Cloud CDN collects access latency data from high volumes of users in real time to dynamically adjust the scheduling policy.

It monitors the linkage quality across the entire network to help allocate the optimal access path.

It supports HTTP/2 and QUIC protocols to reduce data transfer latency and protect data security.

- Diverse experiences in internet businesses

Tencent Cloud CDN leverages Tencent's years of experience in internal business operations

and expertise gained from external business operations on Tencent Cloud.

It supports multiple business models such as bandwidth surge, high IO, and large traffic.

It offers solutions for various scenarios such as emergency event and trend monitoring.

- Flexible billing and project management

Tencent Cloud CDN is pay-as-you-go, and you can change the billing mode based on actual needs to reduce costs.

It supports project-based management for easier use.

- Flexible configurations and various reports

Tencent Cloud CDN allows customization of configuration items such as cache policy to help you flexibly control CDN services.

It provides real-time monitoring and alarms to promptly report problems.

It offers statistical analysis reports and log information to help you make informed business decisions.

2 Tencent Cloud CDN Solution Architecture

2.1 Solution Overview

Tencent Cloud CDN addresses common problems of slow access and low availability of content such as webpages, dynamic content, and streaming media in complex and ever-changing network environments. It has established a CDN cache network consisting of edge and intermediate nodes deployed across regions. Together with a global monitoring system and a comprehensive operations management system, they constitute the Tencent Cloud CDN architecture platform. This "management, content, and monitoring" architecture is as shown in Figure 4:

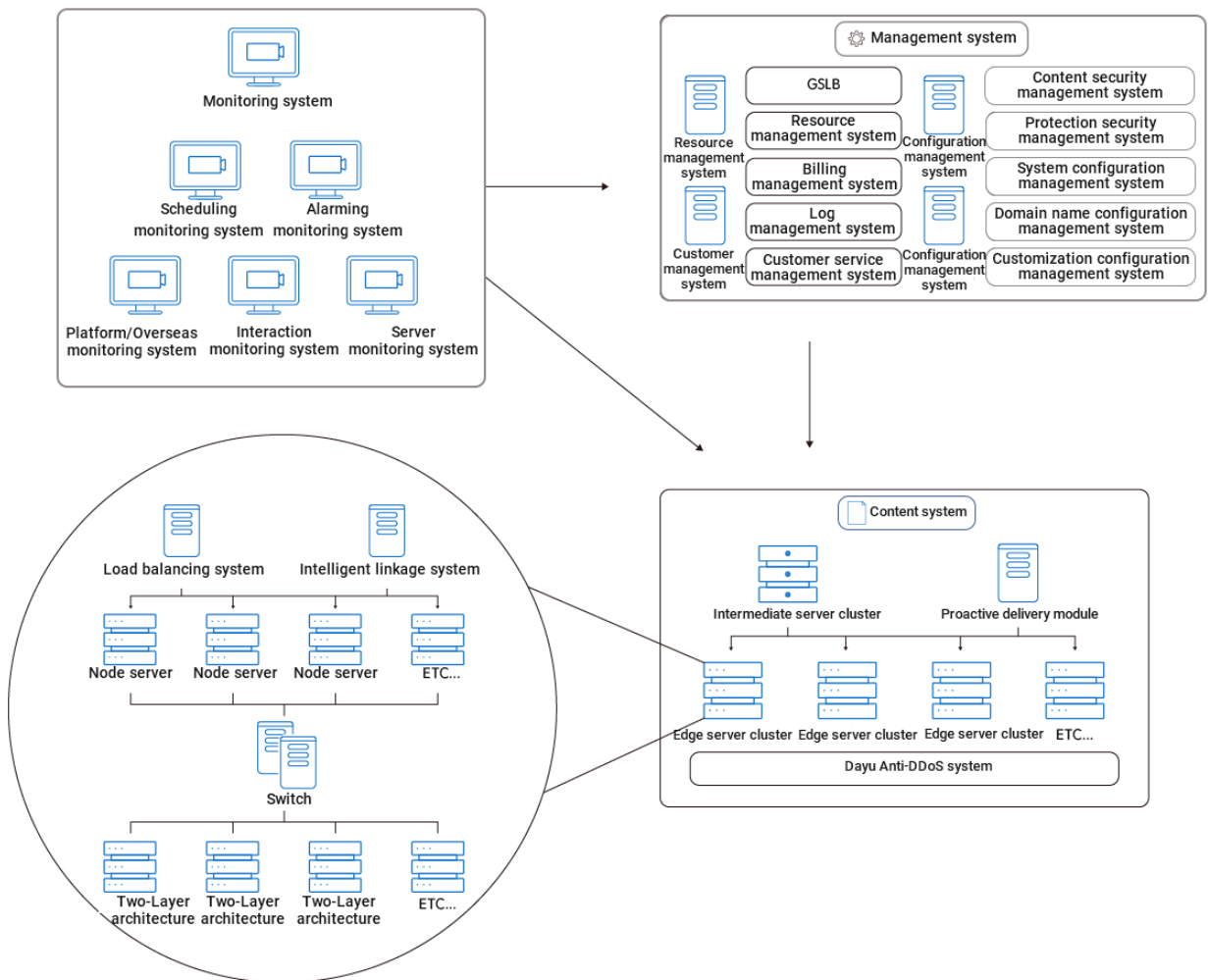


Figure 4. Tencent Cloud CDN platform architecture

2.2 Content System

2.2.1 DiskTank3 Storage Engine

For different file sizes, Tencent Cloud CDN nodes are equipped with different disk types (SSD

or SATA) to perform acceleration and guarantee the hit rate of CDN edge servers. Many commonly used CDN cache components such as Squid, however, cannot perform equally well on both SSD and SATA disks at the same time. This inability to guarantee both I/O performance and acceleration limits the service capabilities of CDN nodes.

As the third-generation storage engine independently developed by Tencent Cloud, DiskTank3 is currently deployed on all servers in the production environment.

2.2.1.1 System-Level Optimization

In general cache management systems, the kernel will use swap memory or reclaim memory when the memory pressure is high, even when the page cache still has free memory. This may lead to problems such as additional CPU overhead, while the file system metadata will also incur I/O overhead. However, the storage engine of Tencent Cloud CDN can manage cached data by itself, reducing file system overhead. As shown in the figure below, DiskTank3 supports direct reads/writes on the disk while bypassing the file system to bring the disk I/O performance into full play. Moreover, the async I/O feature provided by the kernel can be used during reads/writes to unleash the full power of CPU and further improve the processing capabilities of the server.

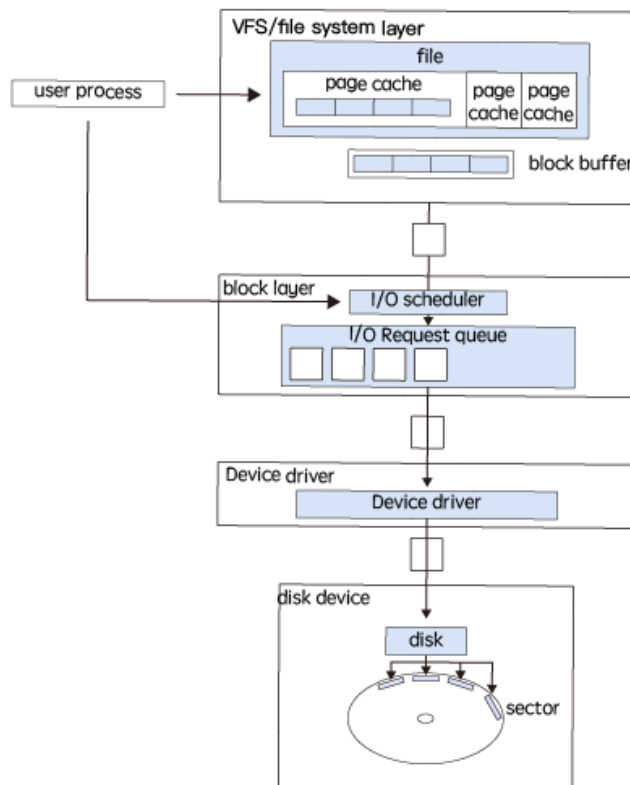


Figure 5. DiskTank3 bypasses the file system

2.2.1.2 Sharded File Storage

Common file systems use contiguous storage. When the file system size reaches the threshold, an old file may need to be deleted before a new file can be received, meaning the entire old file must be deleted from the cache. If the old file size is small, the impact is negligible. However, problem occurs if an old 1-GB file needs to be deleted to store a new 4-KB file. If the large file is requested in a short time, it needs to be obtained through origin-pull or an intermediate server. In DiskTank3, however, all large files are divided into small data shards for storage. As such, only data shards need to be deleted if needed. When the original large file is requested, only the deleted shards need to be obtained again. This solves the problem where the cache file system generates high origin-pull traffic when deleting large files.

In addition, CDN cache system with sharded storage supports variable-length files, which is useful if the origin server only supports HTTP chunked transfer encoding. Because the traditional file system needs to determine the exact file size before storage, it usually receives and caches data in the memory first, determine the file size after receiving the entire file, and then stores it in the cache. In contrast, DiskTank3 can directly write data into the cache, reducing memory and CPU overhead.

2.2.1.3 Metadata Isolation

In many scenarios, metadata is frequently read and written, resulting in I/O overhead. If the metadata is stored in the same disk as the normal file data, access to the latter will be compromised. To solve this problem, DiskTank3 can store metadata and file data separately. Specifically, metadata can be stored in SSD disks with high I/O capabilities and file data can be stored separately in data disks, avoiding metadata I/O overhead.

2.2.1.4 Single-Disk Disaster Recovery

When you operate a large number of servers, disks may occasionally be corrupted. If you wait for the server to trigger an alarm and then perform manual intervention, the response time is long and business operation on the server may be affected. The new version of DiskTank3 can automatically remove corrupted disks from the cache without interrupting business operation. After the corrupted disks are removed, the cache will be hashed again and allocated to the remaining disks without affecting normal file access. This feature can work together with sharded file storage to minimize the impact of disk corruption.

2.2.2 Two-Layer Cache Architecture

The CDN edge server architecture generally has only one layer. To balance CPU cost and efficiency with the cost to mount disks (SSD/HDD), a single server in a cluster often adopts the one-layer configuration. If data cannot be cached on edge servers, requests will be forwarded to the origin server or intermediate server to get the requested resources. As such, the node hit rate and cache utilization efficiency are low.

To address this issue, Tencent Cloud CDN uses a two-layer cache architecture in edge servers and intermediate servers, meaning cache servers in the cluster are divided into two layers:

- Cache servers at layer 1 mount SSD disks to balance speed and processing accuracy.
- Cache servers at layer 2 mount a large number of HDD disks to store massive amount of data. If cache miss occurs at layer 1, data request will be sent to layer 2.
- Only when cache miss occurs at both layers will data request be sent to intermediate servers.

This solution greatly improves the hit rate of edge servers, while ensuring fast processing and reads of high volumes of small files. It also has a similar capability as that of a nearby intermediate server, enabling it to meet the storage needs of large files, achieving an optimal balance between storage and efficiency.

2.2.3 NWS Node Architecture

All edge servers of Tencent Cloud CDN use Next Generation Webserver (NWS) to deliver the optimal service performance. A high-performance HTTP server independently developed by Tencent, NWS has optimized features and performance compared to commonly used servers such as Nginx.

Instead of the multiprocessing mechanism, NWS uses a single-process, multi-thread and event-driven model where each connection is processed by an independent thread. This not only reduces the overhead caused by context switch, but also improves the efficiency and performance of scheduling between processes.

2.2.4 SCDN Platform

Secure Content Delivery Network (SCDN) is a distributed anti-attack security platform based on thousands of Tencent Cloud CDN nodes deployed around the world. It can provide professional DDoS protection services for various business scenarios and attacks such as CC attacks, web intrusions and more. Through dynamic scheduling, it can effectively utilize the redundant bandwidth and protection capability of nodes over the entire Tencent Cloud network to ensure business continuity in case of high attack traffic. In addition, it can prevent website information from being intruded or tampered with by hackers. SCDN provides dedicated policies for websites with high security vulnerabilities and captures packets in real time to formulate appropriate policies for special types of attack. Two sub-products can be configured at the same time based on industry types and actual scenarios:

- The SCDN platform is a value-added service launched by Tencent Cloud CDN to protect businesses, such as games, finance, and websites, that have service unavailability due to heavy DDoS attack traffic. Backed by 2Tbps protection bandwidth and 21 BGP paths, SCDN platform allows you to configure DDoS protection for your websites, divert traffic to the SCDN protection system, and cleanse attack traffic. This ensures that the origin server can provide services reliably and stably, eliminating your concerns over DDoS attacks while accelerating access.

Tencent Cloud CDN features automatic switch to the protection platform. When it detects that

your domain name is under attack, it will automatically switch the domain name to the SCDN platform to ensure business continuity and security, and switch back after the attack ends. This helps minimize the impact caused by sudden attacks on your businesses and platforms while ensuring high-quality acceleration.

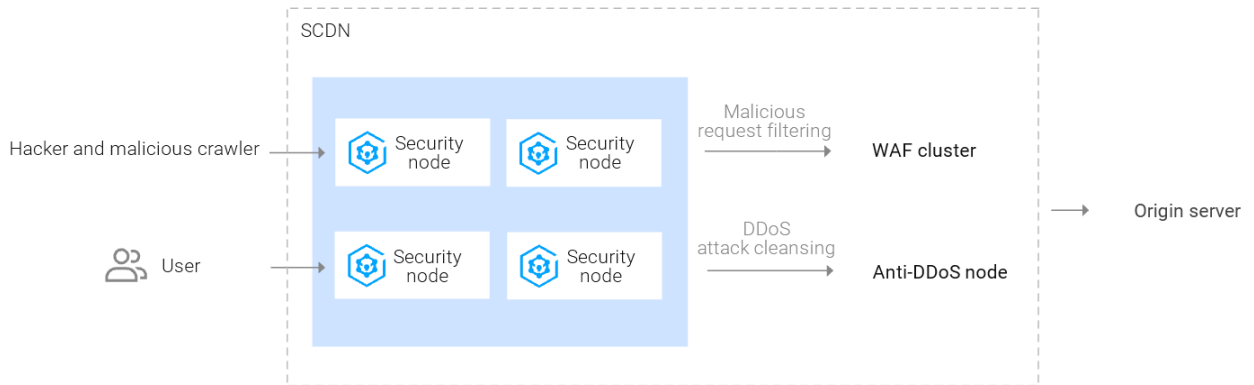


Figure 6. SCDN system architecture

The blocking system of SCDN uses the SynX anti-DDoS security modules developed by Tencent Cloud for targeted identification and analysis of DDoS attack packets based on attack characteristics. It can efficiently and quickly filter malicious attack packets for cleansing and allow normal traffic to pass. SynX directly implements filtering on elastic network interface (ENI) hooks, significantly improving the packet processing capability of the system.

2.3 Monitoring System

2.3.1 Multidimensional Monitoring System

Multidimensional monitoring is a comprehensive monitoring method that evaluates the status of servers on CDN nodes from all aspects and then adjust the traffic distribution across the entire network to ensure the optimal performance of your business. At the underlying layer, based on the TLinux operating system developed by Tencent Cloud, it can comprehensively monitor basic hardware data. Based on the high-performance web server NWS developed by Tencent Cloud, it can collect current quality access data at the application layer in real time. Based on testing of external systems (dedicated automated testing points and QQ clients), it can accurately collect nationwide network access quality data. By aggregating the monitoring data, Tencent Cloud CDN can adjust the service quality in real time to provide optimal services.

Multidimensional monitoring method:

- Single node health monitoring: data on CPU, I/O, and network status is collected on the network layer and hardware layer, and last-mile metrics are monitored from the nodes to end users.
- Multidimensional monitoring of the CDN platform: NWS architecture logs are monitored in real time to ensure the quality of CDN operations. It proactively processes an alarm rather

than responding to it passively.

- Connected test operation system: Tencent Cloud CDN has many testing terminals and methods. QQ and WeChat, two of the largest businesses operated by Tencent, use the same test platform. This massive amount of network data forms a "traffic map" that covers the entire network. Backed by the most extensive network coverage and highest number of users among all CDN and cloud service providers, Tencent Cloud CDN can instantly collect network quality data to adjust the node coverage.
- For hardware services, OPS of global CDN nodes relies more on accurate and fast monitoring. The stringent restrictions of the dedicated global monitoring system ensure that potential problems in the node cluster can be filtered automatically and the OPS personnel will be notified immediately.
- In general, the quality of cross-border network is worse than that of domestic network. Tencent Cloud CDN breaks the bottleneck of cross-border data transfer by fine-tuning the BBR congestion control algorithm. However, a complete monitoring system is still needed to ensure the normal operation of services, especially dynamic acceleration. When the monitoring system detects latency in a cross-border connection, it will automatically recalculate and switch connections to ensure the normal operation of the global network.

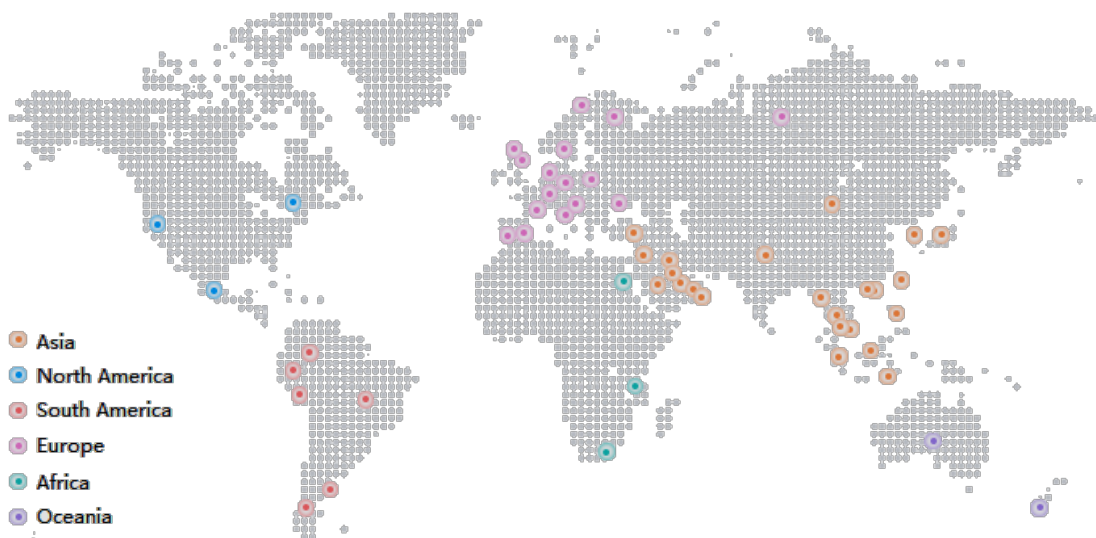


Figure 7. Global distribution of Tencent Cloud CDN nodes as of 2020

2.4 Management System

The management system of Tencent Cloud consists of four parts, which are mainly categorized by service type: resource management, security management, configuration management, and customer management. From the management of individual servers, domain names and platform configurations, to customer-oriented integrated management, a comprehensive management system

is in place. In addition, as one of the most important systems, security is separately managed and configured. The following analyzes key technologies in each system:

2.4.1 GSLB

The resource management system consists of Global Server Load Balancing (GSLB) for resource scheduling and the resource management system for resource coordination and management. Currently, the mainstream CDN scheduling method is to allocate nearby nodes based on DNS address. Due to the complexity of the Chinese network and ISPs' unstandardized IP libraries as well as egress routers, scheduling error often occurs.

GSLB can monitor the status of servers on Tencent Cloud CDN nodes and user latency in real time, while syncing the monitoring data to the scheduling decision-making module. The module analyzes the monitoring results and syncs information of the optimal node to DNS for final resolution as shown in Figure :

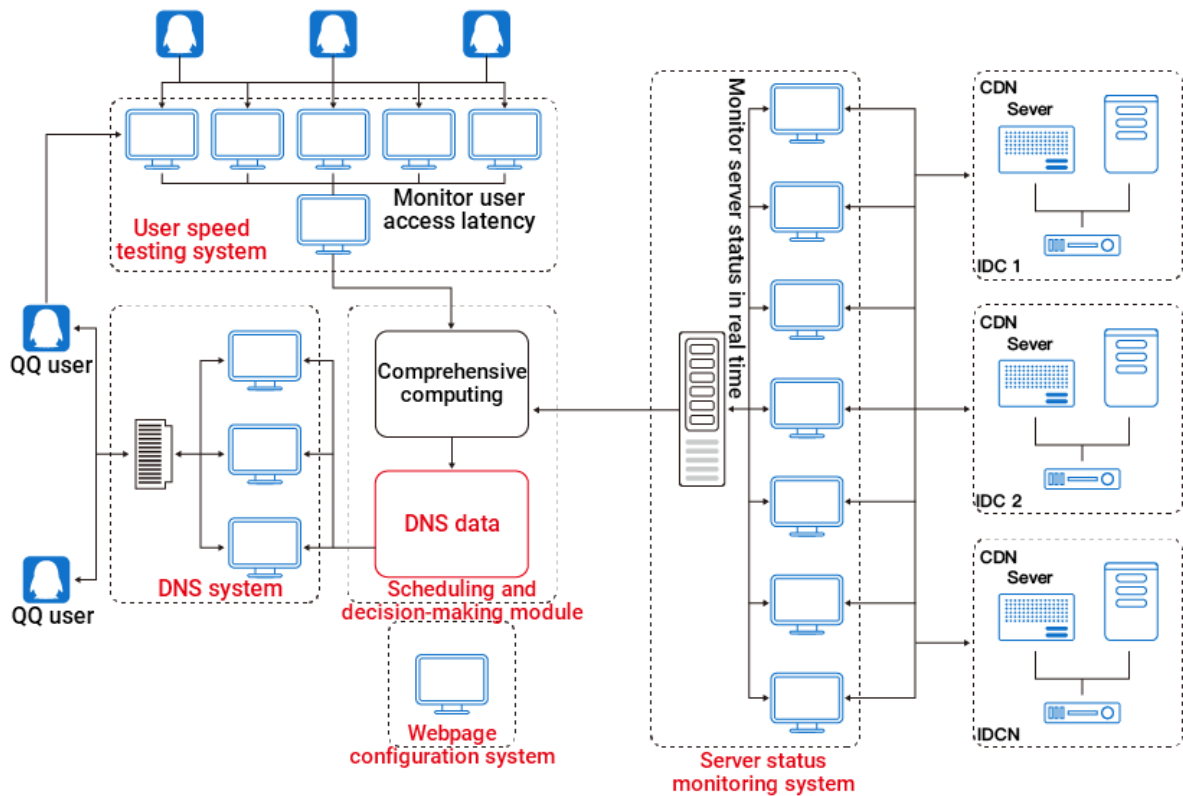


Figure 8. GSLB architecture

This solution not only supports access using the standard DNS protocol, but also provides the HttpDNS access solution. HttpDNS can guarantee the scheduling quality and prevent DNS hijacking, ensuring scheduling accuracy and security.

2.4.2 Configuration Management System

The configuration management system consists of the configuration management for the system,

domain name, and customization.

- System configuration management is used to forward CDN system configuration items and maintain system operation as well as certain global policies.
- Domain name configuration management: Tencent Cloud CDN uses distributed domain name configuration management servers to quickly distribute configurations level by level based on CDN principles. Domain name configurations can be forwarded and deployed to the entire network within 5 minutes. The high efficiency ensures that configurations can take effect instantly.
- Customization configuration management: Tencent Cloud CDN can meet your custom needs, and manage custom and standard feature configurations separately to prevent potential impact. This ensures that standard features operate normally while your custom needs are also met.

3 Tencent Cloud CDN Platform Solution

Based on the accelerated content, Tencent Cloud CDN platform is divided into webpage acceleration, download acceleration, and video on demand (VOD) acceleration platforms. Despite their differences, these platforms share some common features.

3.1 Common Platform Features

3.1.1 HTTPS Optimization

Since early 2015, some mainstream websites have adopted HTTPS to improve website security and prevent problems such as traffic hijacking and product placement. Baidu and Alibaba have implemented HTTPS for their entire sites. According to Google, however, the single-core performance of a CPU is about 1,500 connections per second (CPS) for a 1,024-bit certificate. As asymmetric encryption in SSL encryption consumes high CPU usage, the cost and pressure to convert the origin server are high. For a 2,048-bit certificate, CPS for the CPU goes down to 300–800. Compared with HTTP requests, the performance is 6–8 times lower, which is a significant reduction. To convert the website without compromising its performance, you need to invest heavily in hardware and labor.

Using an SSL certificate means the HTTPS protocol is used, but security is not fully guaranteed at the HTTPS level, even if the purchased certificate and service used meet the requirements of ATS protocol recommended by Apple:

- The server must support TLS 1.2 or above;
- The password must support forward secrecy;
- The certificate must use the Secure Hash Algorithm 2 (SHA-2) with a digest length of at least 256 bits;

Still, many factors may compromise HTTPS security during actual use. The well-known website, ssllabs.com, provides an evaluation solution for HTTPS security.

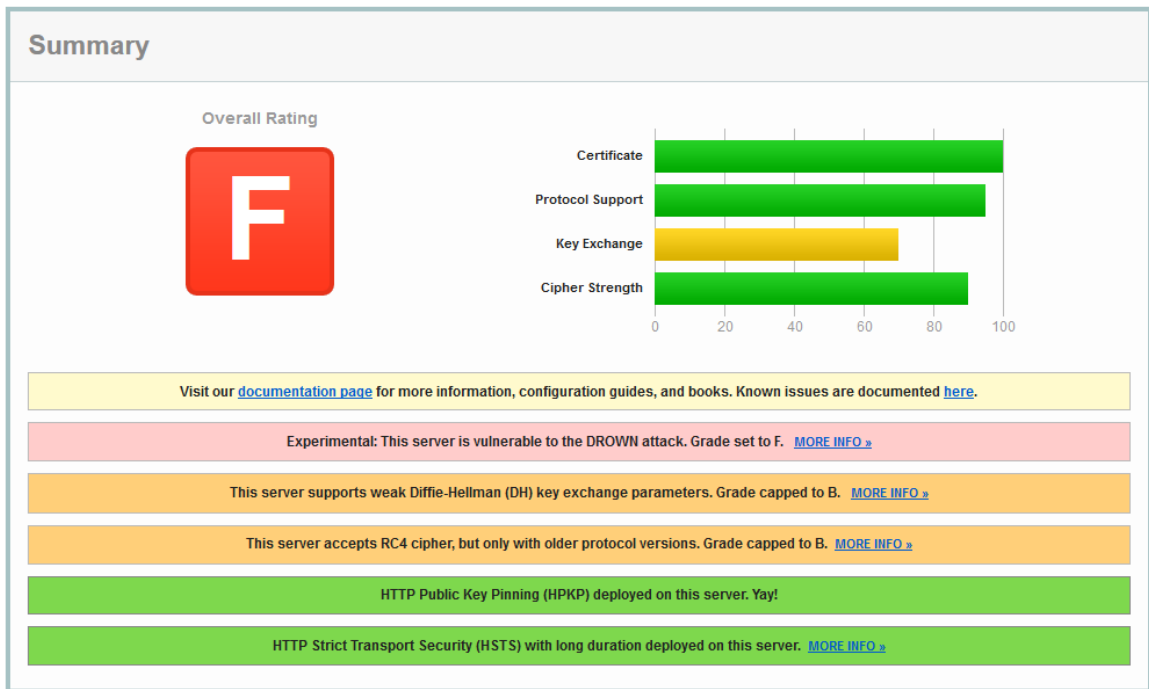


Figure 9. SSL security test result of a website

As shown in Figure 9, while the purchased certificate meets the compliance requirements, SSL security level of the CDN platform is determined by the overall security policies and measures.

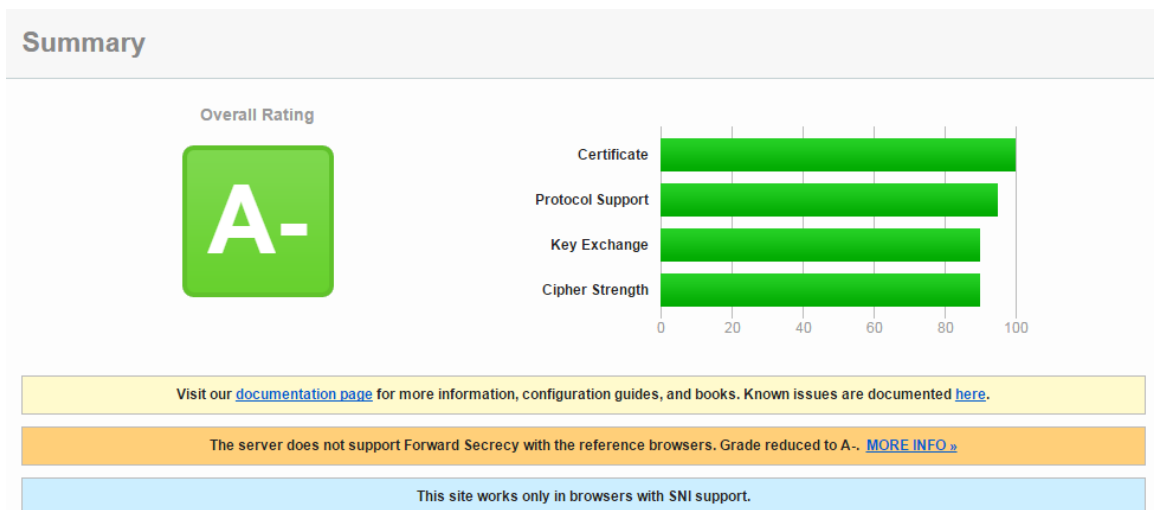


Figure 10. Tencent Cloud SSL security test result

As shown in Figure 10, the entire Tencent Cloud CDN platform achieves A-level protection.

3.1.1.1 Certificate Solution

Tencent Cloud CDN has three HTTPS certificate solutions:

- We offer a flexible solution in which the content is encrypted from client to CDN with a secure HTTPS connection using CDN's certificate. Content is returned from CDN to the origin server via HTTP with no modification needed on the origin server.
- If your own certificate has been deployed on the origin server but should be kept private, content is encrypted from client to CDN with a secure HTTPS connection using CDN's certificate. Content is returned from CDN to the origin server via HTTP with no certificate information needed from the origin server.
- If your own certificate has been deployed on the origin server and is open to other users, end-to-end secure acceleration can be achieved from client to origin server.

You can choose the most appropriate solution based on actual needs.

3.1.1.2 Acceleration Technology

3.1.1.2.1 Fast SNI Deployment

In the legacy CDN deployment method (which is still used by many CDN platforms), because domain name information is not passed during SSL handshake, the server usually returns the first available certificate in the configuration. If multiple certificates are needed, you have to configure different SSL ports or add IP addresses, or purchase "multi-domain SSL certificate" or "wildcard certificate" that is very expensive. To deploy a certificate on CDN nodes, you need to bind the certificate the same way you bind IPs. This process is time-consuming, while modification and adjustment of the certificate coverage are complicated.

Server Name Indication (SNI) is a technology used to optimize SSL/TLS and is enabled in SSLv3/TLSv1. It allows the client to specify the hostname it is trying to reach when initiating an SSL handshake request, so the server can switch to the correct domain and return the corresponding certificate. This helps CDN system simplify the process of certificate deployment and adjustment, and reduce the time for adjusting resource coverage and reporting problems.

3.1.1.2.2 AES-NI

AES-GCM is a popular symmetric encryption algorithm. It solves the security vulnerability of symmetric encryption, but also causes performance problems. To cope with this, Intel has launched Advanced Encryption Standard New Instruction (AES-NI), an extension to the x86 instruction set architecture for microprocessors from Intel and AMD. It can improve AES performance at the hardware level, and is now widely supported by CPUs on servers and PCs.

As shown in Figure 11, according to Intel, server performance improves by 5–8 times after AES-

NI is enabled.

AES-128-GCM performance with AES-NI enabled and disabled

AES-128-GCM	AES-NI Enabled	AES-NI Disabled
16 Bytes	404141.90	72163.37
64 Bytes	994097.92	96301.44
256 Bytes	1421557.93	215897.60
1024 Bytes	1515602.94	231689.22
8192 Bytes	1529460.05	235077.63

Figure 11. AES-NI effect comparison

3.1.1.2.3 False Start

TLS False Start means the client sends application data (such as HTTP request) when sending "Change Cipher Spec Finished" (second stage of SSL handshake), and the server directly returns the application data (such as HTTP response) when TLS handshake is completed. In this way, application data can be sent before the handshake is fully completed, as shown in Figure 12.

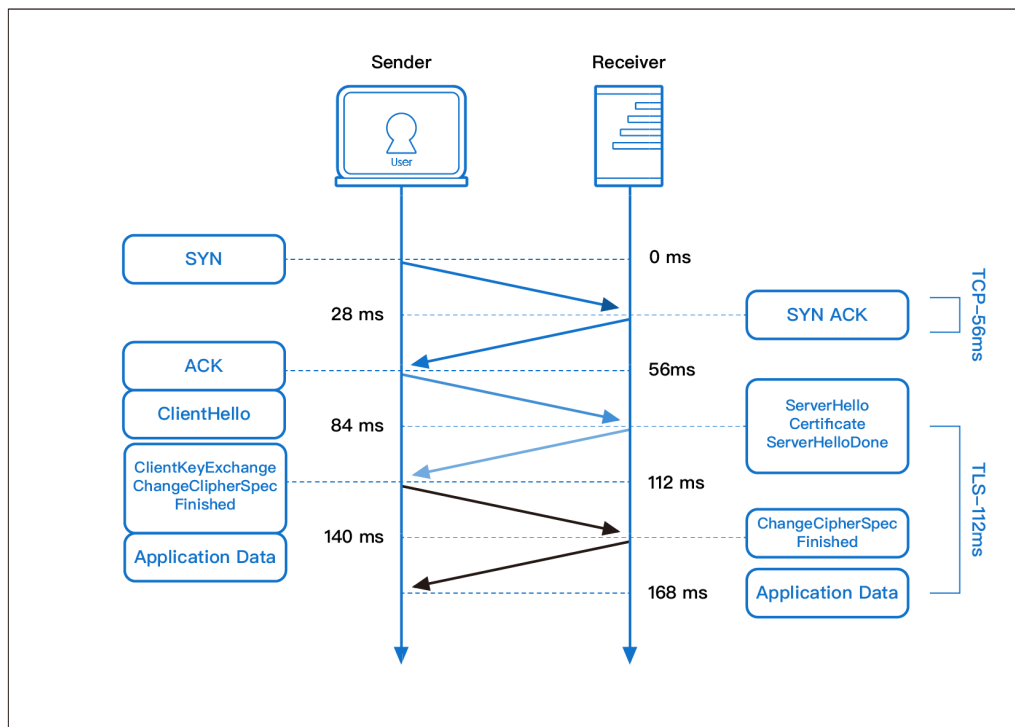


Figure 12. False Start flowchart

After False Start is enabled, only one RTT is required during TLS handshake to start application data transfer. False Start means the client sends encrypted application data in advance, delivering a faster 1-RTT TLS handshake with no modification required for TLS protocol. Currently, False Start is enabled in most browsers by default.

3.1.1.2.4 OCSP Stapling

Sometimes, the CA may need to revoke certain certificates. The certificate holder usually has two methods to check whether the certificate is invalid: Certificate Revocation List (CRL) and Online Certificate Status Protocol (OCSP).

However, some clients will query the OCSP API in real time during TLS handshake and block subsequent process before getting the result, significantly compromising the performance. With OCSP stapling, however, the server will append the OCSP response signed by the CA to the certificate chain, eliminating the need for the browser to validate the certificate. Thus, the server has a faster network and it is easier to get the OCSP response, which can also be cached.

The OCSP response is encrypted and cannot be forged. OCSP stapling improves both handshake efficiency and security.

3.1.1.2.5 HTTP/2 Acceleration

Hypertext Transfer Protocol (HTTP) standard has recently been upgraded. HTTP/2 was approved in May 2015, and is currently used in many web browsers and servers (including Nginx Plus and open-source Nginx). About two-thirds of browsers already support HTTP/2.

Key features of HTTP/2 are derived from SPDY, including but not limited to:

- HTTP/2 is a binary (not plain text) protocol, which is simpler and more efficient.
- HTTP/2 uses one multiplexed connection per domain instead of one connection per file.
- HPACK compression format is used for the request header, instead of gzip used in SPDY.
- Tencent Cloud CDN supports the HTTP/2 protocol, and the acceleration result is significant for HTTPS.

3.1.2 Hotlink Protection

Hotlinking is a malicious behavior where a service provider does not provide its own content but uses plagiarized content belonging to others, while bypassing profit-making end-user interfaces (such as ads) through technical means. While the malicious service provider gains views and clicks by providing no or few resources, others could have their revenue sources undermined.

To meet the hotlink protection requirements of most users, Tencent Cloud CDN provides four hotlink protection methods:

3.1.2.1 Referrer Hotlink Protection

3.1.2.1.1 How It Works

It filters the request content through the HTTP referer and returns the corresponding information.

HTTP referer is a part of an HTTP request header. When the browser sends a request to the web server, referer will be carried so the server knows which link does the request come from in order to process the information.

3.1.2.1.2 Notes

- Referer blacklist and whitelist are mutually exclusive;
- If both the hotlink protection blacklist and whitelist are empty, hotlink protection is disabled and the rules do not take effect;
- Up to 200 entries can be added for hotlink protection;
- Hotlink protection rules are matched as follows:
 - Prefix match of domain name/IP rule is supported.
 - Domain name prefix wildcard is supported.
 - By default, no rules will be matched If the referer is empty (i.e., when user accesses the domain name directly in a browser).
 - If the referer whitelist contains an empty referer (i.e., when the resource URL is accessed directly via the address bar of a browser) when you select it, CDN will return the resource information normally.

3.1.2.2 User-Agent Hotlink Protection

User-Agent header contains information of the client (usually a browser) that initiates the original request. User-Agent information varies by client. Similar to referer, Tencent Cloud CDN supports rejecting or allowing HTTP requests whose User-Agent information in their headers meets specific User-Agent rules (regex match supported). However, the security of User-Agent hotlink protection is relatively low because User-Agent can be easily forged.

Use Cases:

1) If you want that the webpage or resource can be accessed only by IE but not Chrome, you can configure User-Agent hotlink protection.

2) If you have a dedicated download tool for content download that uses its own custom name in User-Agent, and you want that your content can only be downloaded by requests that meet this rule, then you can configure User-Agent hotlink protection.

3.1.2.3 URL Hotlink Protection

URL hotlink protection is used to protect your video content. With this feature, you can generate real-time playback links that can be accessed via players or URLs and are valid for a specified time period to effectively protect video content. As the hotlink protection uses dynamic links, the original player code and file URLs become invalid when this feature is enabled.

3.1.2.3.1 How It Works

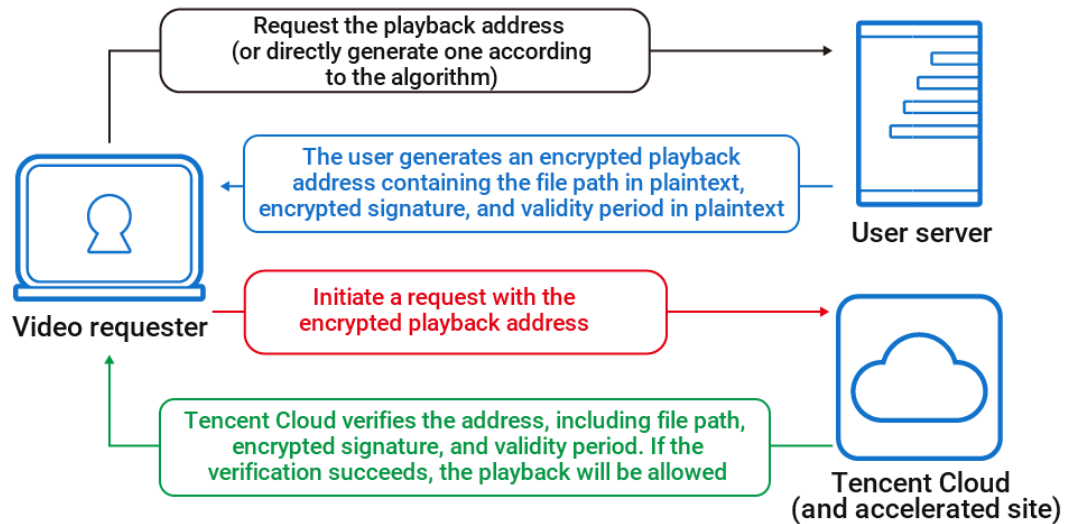


Figure 13. URL hotlink protection flowchart

1. Enable the service and get the key: after you apply for this feature, Tencent Cloud will send you the key, which should be kept private.
2. Generate the encrypted playback address: when the file or video requester initiates a request, the access address can be generated automatically based on the public-key algorithm or calculated by the server. The generated address is in the format of "http://www.test.com/folder/vodfile.mp4?sign=abc123dsaadsasdads&t=55bb9b80". The fields are generated as follows: sign=abc123dsaadsasdads (example): it is the verification string and the algorithm is:

"sign = MD5(KEY+ path +t)" with lower cases:

 - a) KEY: the key assigned by Tencent Cloud CDN in step 1;
 - b) path: the file path in the URL excluding characters after "?". It is "/folder/vodfile.mp4" in the example;
 - c) t: the timestamp for the validity period of this link. In this example, it is "4d024e80", which is a hex-based UNIX timestamp;
3. The file or video requester generates a link based on the calculation result above and initiates the request to Tencent Cloud and Tencent Cloud CDN;
4. Tencent Cloud verifies the link and determines whether to play back the video: it will first check whether the validity period has passed based on the timestamp in plaintext. If not, it

will verify whether the verification string is correct. Only when both conditions are met can the video be played or the file request be allowed.

3.2 Webpage Acceleration Platform

The webpage acceleration platform of Tencent Cloud CDN is optimized for webpage images by taking into account the fact that there are many dynamically requested and frequently updated small image files on webpages. It also provides acceleration policies customized for dynamic and static content on webpages, such as dynamic acceleration and lossless image compression. In addition, it supports quick content delivery and intelligent route origin-pull to reduce the latency end users may experience when accessing different content. The effects are as follows:

- Client: access speed and quality (availability) are improved.
- Origin server: origin-pull traffic and pressure as well as investment on origin servers are reduced.
- Service: service quality remains stable even during possible bandwidth surges to ensure business continuity.
- OPS: malicious intrusions and high-traffic attacks are blocked to guarantee secure OPS of origin servers. In addition, complete and accurate user logs as well as multidimensional statistical analysis are provided to keep OPS personnel informed on system conditions at any time.

3.2.1 Applicable Scenarios

The webpage acceleration platform is mainly used to solve slow access and low availability of various PC/mobile websites and mobile apps in the current complex network environment, especially the mobile internet. It is suitable for content transferred over HTTPS/HTTP, and can successfully accelerate pages and applications with a lot of images and text content, such as portal websites, ecommerce platforms, games, reading services, social media, government websites, and office systems.

3.2.2 Platform Architecture

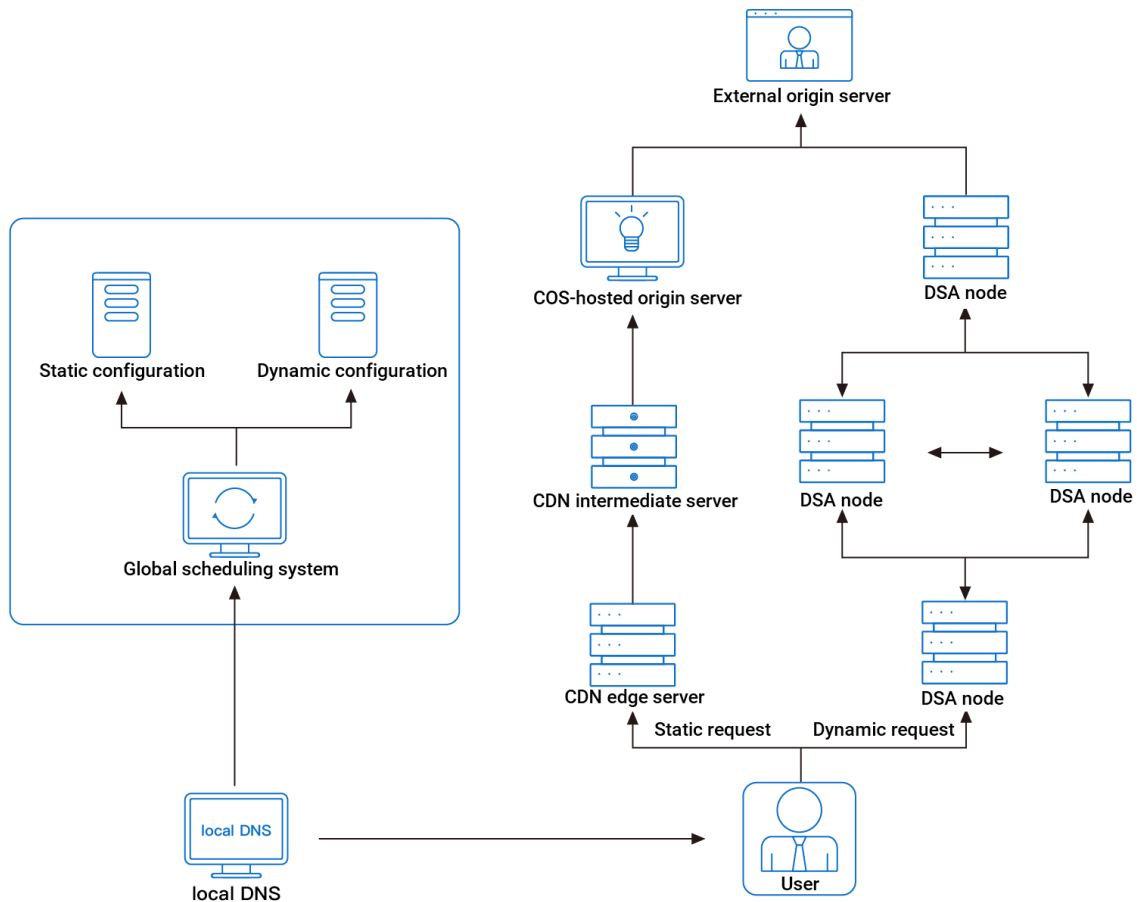


Figure 14. Webpage acceleration platform architecture

As shown in Figure 14, the architecture of the webpage acceleration platform mainly consists of the following parts:

1. **Global node scheduling:** the global scheduling system identifies the access location of the end user based on the LocalDNS access IP. It can also determine the service quality of QQ and WeChat clients accessed via the entire network, and the current load and health of nodes based on the GLSB decision-making module. It can then assign optimal nodes to end users.
2. **Clusters of edge servers and intermediate servers:** as described in section [2.2.2 Two-Layer Cache Architecture](#), for a static file, if the cache in the two-layer cache cluster of edge servers cannot be hit, the request will be forwarded to a CDN intermediate server. If the cache is hit on the intermediate server, the cached content will be returned to the client through an edge server, which will also cache the content. Otherwise, the request will be forwarded to the origin server, and both the intermediate and edge servers will receive and cache the content.

For more information on dynamic requests, please see section 3.2.3.1. DSA Dynamic Acceleration below.

3. COS-hosted origin server: Cloud Object Storage (COS) is a highly available, stable, and secure cloud storage service for enterprises and individual users. The origin server can put unstructured data in any amount and form into COS for easy management and processing. Data content can be stored on origin server hosted by COS both actively and passively. This helps significantly reduce the pressure on origin servers. (For more information on COS, please visit <https://cloud.tencent.com/product/cos>.)

3.2.3 Platform Features

In addition to the general CDN cache features described above, the webpage acceleration platform also has the following features for different content types and use cases:

3.2.3.1 DSA Dynamic Acceleration

Both mobile apps and websites have a large number of dynamic requests. In scenarios such as securities trading, most requests are dynamic. As shown in Figure 15, when loading websites and pages, dynamic resources in the resource list need to be loaded first before static resources are requested one by one. The latency of dynamic requests may slow down the loading speed of the entire page and compromise user experience.

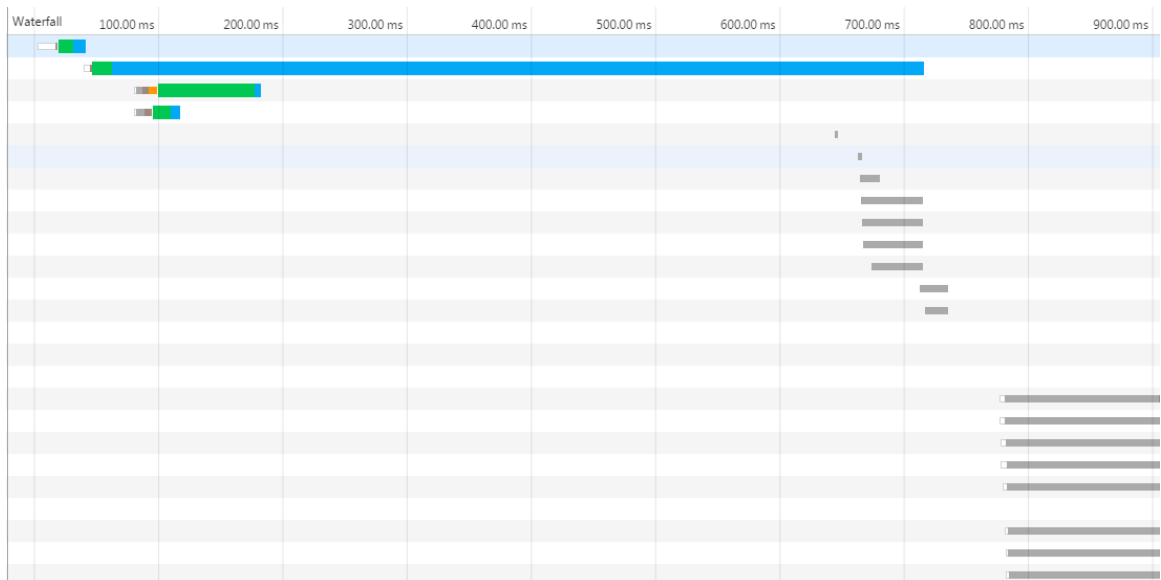


Figure 15. Effect of dynamic requests on loading speed

Despite urgent demands for acceleration of dynamic requests, bottlenecks and problems remain to be solved.

1. As the number of internet users grows, network traffic also soars. With the rapid development of smart phones and PCs, users have higher demands for high-definition images and videos, further causing an increase in network traffic. However, the construction speed of backbone network cannot meet user needs, which is also limited by poor ISP routing policies and cross-ISP bandwidth that may compromise transfer speed and cause network congestion. The key question facing dynamic request acceleration is how to avoid congested network and locate

the optimal origin-pull path.

2. The traditional TCP protocol is complicated and still under development. Although optimization has been proposed and implemented (for example, RFC 2581 TCP Congestion Control is an important RFC on TCP in recent years), basic operations of the TCP protocol described in RFC793 remains unchanged. Published in 1981, it is no longer suitable for the current complex network environment, significantly affecting the acceleration of dynamic requests.
3. CDN aims to solve the problem with dynamic acceleration, but common CDN architecture can be counterproductive. The strict request path from edge server to intermediate server in CDN may lengthen the origin-pull path and prolong the request time.

Tencent Cloud CDN has launched Dynamic Site Accelerator (DSA) to solve the problems facing dynamic request acceleration from various aspects, such as network architecture, communication technology, and hardware architecture.

3.2.3.1.1 Network Structure

In general, access and storage of static resources follow a strict hierarchical chain structure. User request will access edge server first and be forwarded to intermediate server only if the requested resource is not cached on the edge server. This hierarchical structure also applies to edge/intermediate server and COS-hosted origin server and COS-hosted origin server.

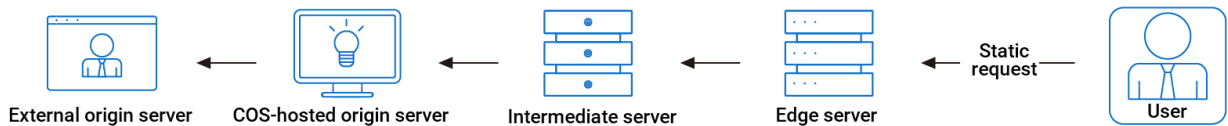


Figure 16. Static content origin-pull process

In DSA, however, there is no strict node hierarchy when user requests access the dynamic acceleration network through nodes assigned by the GSLB scheduling system. The request will be forwarded to the fastest transfer route from access to origin-pull. Each DSA node stores its connection information with other nodes, such as latency and packet loss rate, and assigns different weights to calculate the optimal origin-pull path.

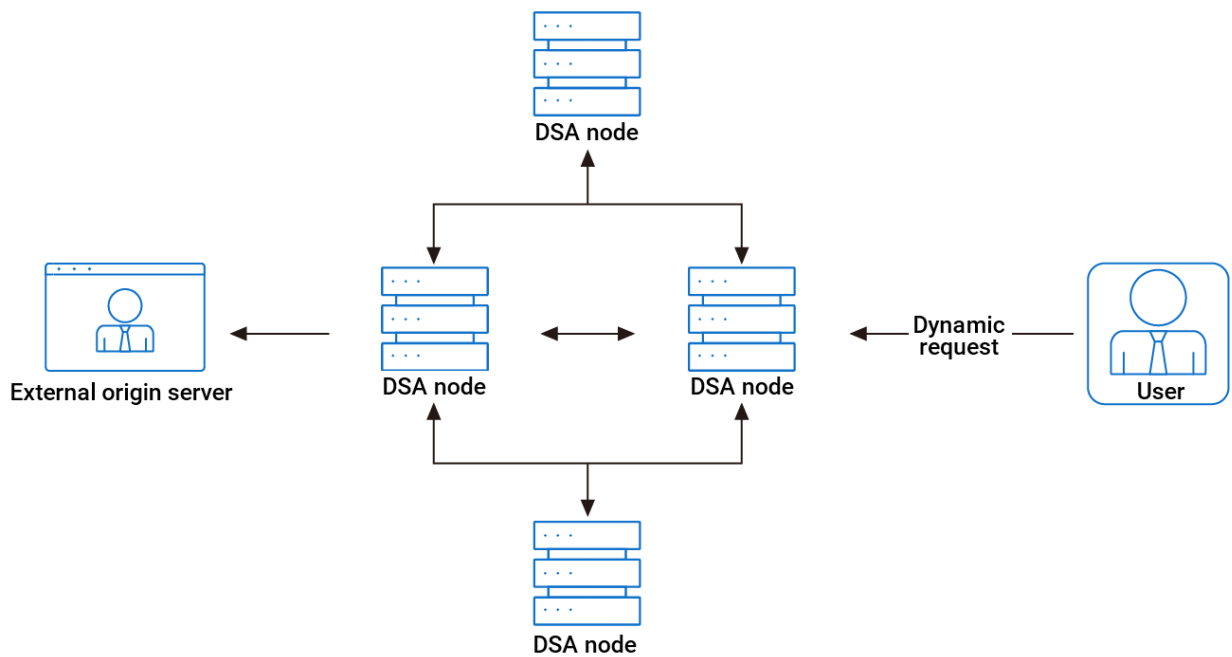


Figure 17. Dynamic acceleration request flowchart

3.2.3.1.2 Private Protocol

TCP protocol, which is theoretically designed for transmission (how many KB of data can be transferred per second) and efficient use of current bandwidth, can be optimized in dynamic transfer. For dynamic requests, however, fast transfer is more important than efficient use of bandwidth. Given this, Tencent Cloud CDN R&D team combines high efficiency of UDP protocol with the reliability of TCP protocol to develop the QCTP transfer protocol. Based on the efficiency of UDP, QCTP also optimizes the protocol from various aspects such as retransmission and ACK mechanisms.

- Congestion control: for TCP, to balance congestion control and traffic control, the initial congestion window size is 1. After the first message is sent and ACK is returned, the congestion window size will increase exponentially, which is the slow start algorithm currently used by TCP. When the congestion window size reaches the limit, it will start an additive increase. This allows the window size to increase slowly to the optimal value while avoiding network congestion caused by the increase speed. When packet loss occurs, however, the congestion window size decreases significantly and enters the exponential increase phase again. For more information, please see Figure 19 **错误!未找到引用源。**

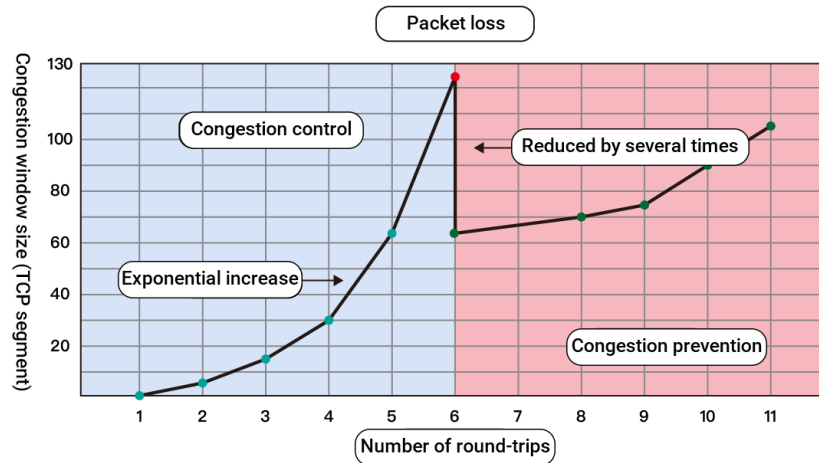


Figure 18. Traditional TCP congestion control

This method avoids network transfer problems caused by packet loss but compromises transfer efficiency, which is not suitable for dynamic content transfer. In contrast, QCTP protocol can dynamically detect the current network conditions between two nodes, and adjust the congestion window size based on factors such as the size of data sent and size of remaining content, which is more efficient than TCP.

- Retransmission algorithm: in TCP, when packet loss occurs, data on and after the lost packet will all be retransmitted, while QCTP protocol only retransmits the packet that is actually lost. QCTP also redefines timeout of the TCP algorithm by anticipating possible timeouts in advance and proactively performing retransmissions. While this may incur waste, QCTP ensures faster and more efficient transmissions. Compared to TCP, DSA improves the transmission efficiency by over 30%, despite higher consumption of bandwidth between internal nodes.

3.2.3.2 Pornographic Image Detection

Based on Tencent YouTu's deep learning technology for pornography detection, this feature can efficiently and accurately detect pornographic images. It helps websites and application developers overcome the challenge of pornography detection and eliminates the need for manual review, improving efficiency and reducing cost while ensuring a healthy network environment.

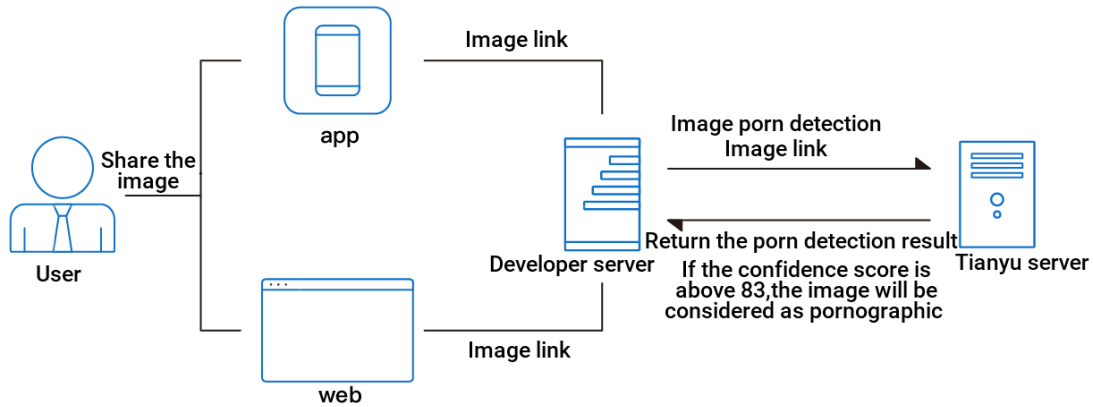


Figure 19. Pornographic image detection system architecture

Pornographic image recognition, sexy image recognition, and automatic pornographic image detection are provided:

- **Pornographic image recognition:** images provided by developers are automatically downloaded and checked for pornographic information, while confidence scores will be returned. Images with scores above 83 are suspected pornographic images. Developers can track and analyze user behavior related to these images.
- **Sexy image recognition:** images provided by developers are checked for sexy information. If images have scores above 83, developers can track and analyze user behavior related to these images. This feature is free of charge.
- **Automatic pornographic image detection** is backed by Tencent YouTu's DeepEye active porn recognition engine to analyze images and return confidence scores. Trained with a large number of image samples from Tencent's social media services, the algorithm can achieve a recognition accuracy of above 99.9%, which far exceeds that of humans and can replace 90% of manpower in practice. To tackle the challenge in the field of automatic image recognition of distinguishing a pornographic image from non-pornographic (sexy) image, DeepEye engine uses graph partition for precise recognition.

3.3 Download Acceleration Platform

Compared with scenarios in section [3.2 Webpage Acceleration Platform](#), the delivery of large files leads to different problems. Downloading large files incurs high bandwidth and fluctuation, while involving a variety of download tools. This brings the following common problems to users: poor download quality, low success rate, untimely software release, unstable bandwidth, poor user experience, hotlinking issues, etc.

3.3.1 Applicable Scenarios

The download acceleration platform is mainly applicable to the following scenarios: various mobile app stores, audio/video file delivery on applications and webpages, game updates/apps/clients,

firmware/hardware/software update downloads, etc. These can be divided into three categories: publisher, download site, and video provider. The acceleration platform provides fast and high-quality download services for content providers with needs for large-scale content delivery by solving problems as listed in Table 1:

Target Customer	Data Size	Network-wide Sync Speed	Bandwidth Requirement	Bandwidth Fluctuation	Response Time	Success Rate	Hotlink Protection Required
Publisher	Small	High	High	Large	Long	High	No
Download Site	Large	Average	Average	Average	Average	Average	Yes
Video Provider	Large	Average	High	Average	Long	High	Yes

Table 1. Applicable scenarios of the download acceleration platform

3.3.2 Platform Architecture

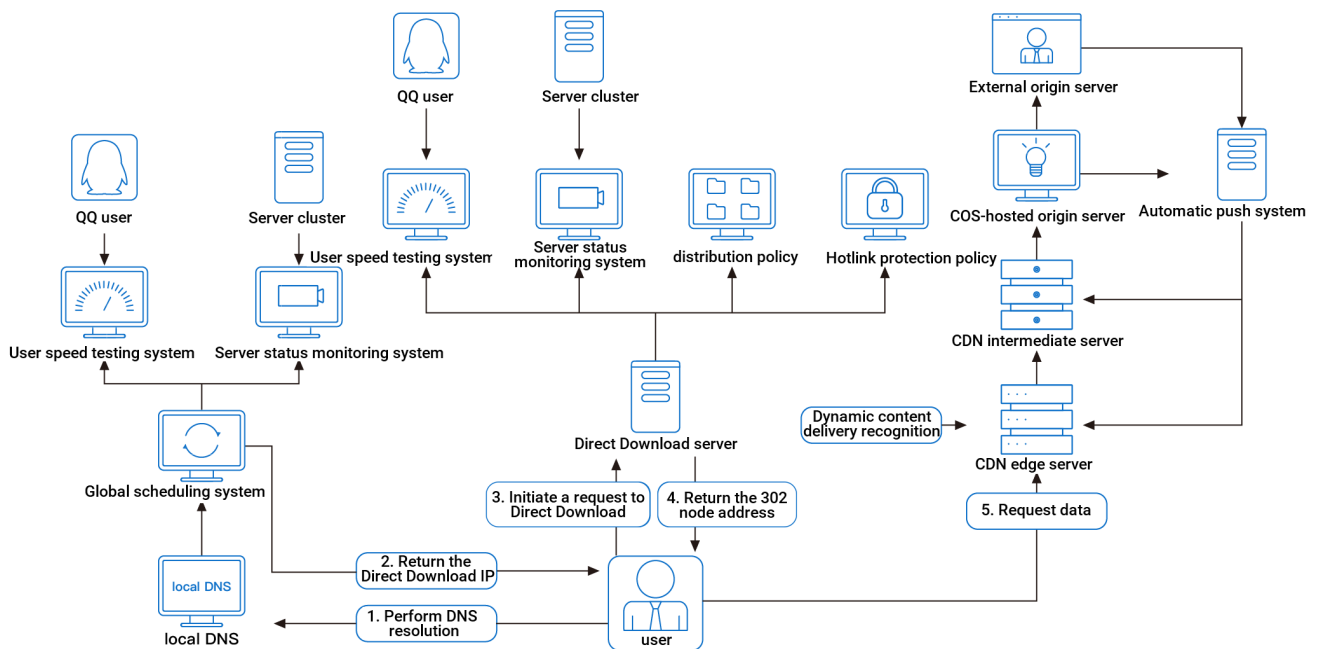


Figure 20. Download acceleration platform architecture

As shown in Figure 20, the architecture of the download acceleration platform is different from section 3.2.2. Webpage Acceleration Platform Architecture: the global scheduling system can be connected with users through dedicated DNS to locate the optimal server resources in real time rather than relying on DNS scheduling and resolution. This fundamentally changes the CDN access process and improves user access. Similar to webpage acceleration platform, the download acceleration

platform processes static files as follows: if the cache in the two-layer cache cluster of edge servers cannot be hit, the request will be forwarded to the CDN intermediate server. If the cache is hit on the intermediate server, the cached content will be returned to the client through an edge server and cached there at the same time. Otherwise, the origin server will be requested for the content, and both the intermediate server and edge server will receive and cache the content. For large files (e.g., mobile phone ROM and game installation packages), however, the bandwidth will surge during passive origin-pull, leading to high pressure on the origin server. In this case, proactive delivery between nodes can effectively control the progress of node caching to reduce bandwidth and pressure on the origin server.

3.3.3 Platform Features

The download acceleration platform of Tencent Cloud CDN is designed to solve the download problems of large and small files. In addition to accelerating file download, it can be integrated with the unique dedicated DNS feature of Tencent Cloud CDN to prevent ISP hijacking and implement global load balancing. Moreover, the dynamic recognition feature changes the traditional top-down delivery mechanism of CDN. Fast file transfer between nodes increases the CDN file prefetch efficiency to the minute, while reducing the pressure on internal nodes and origin servers.

3.3.3.1 Dedicated DNS

Given the complex network environment in China, scheduling to the closest server based on DNS resolution may encounter problems such as scheduling failure and long latency. With dedicated DNS, Tencent Cloud CDN's globally deployed server clusters for 302 redirect can determine the optimal server resources for each request in real time, solving the scheduling problems of small ISPs and improving user access quality. Using dedicated DNS for scheduling has zero latency, and user request can be forwarded to the optimal server node as quickly as possible while bypassing nodes with poor performance or exceptions.

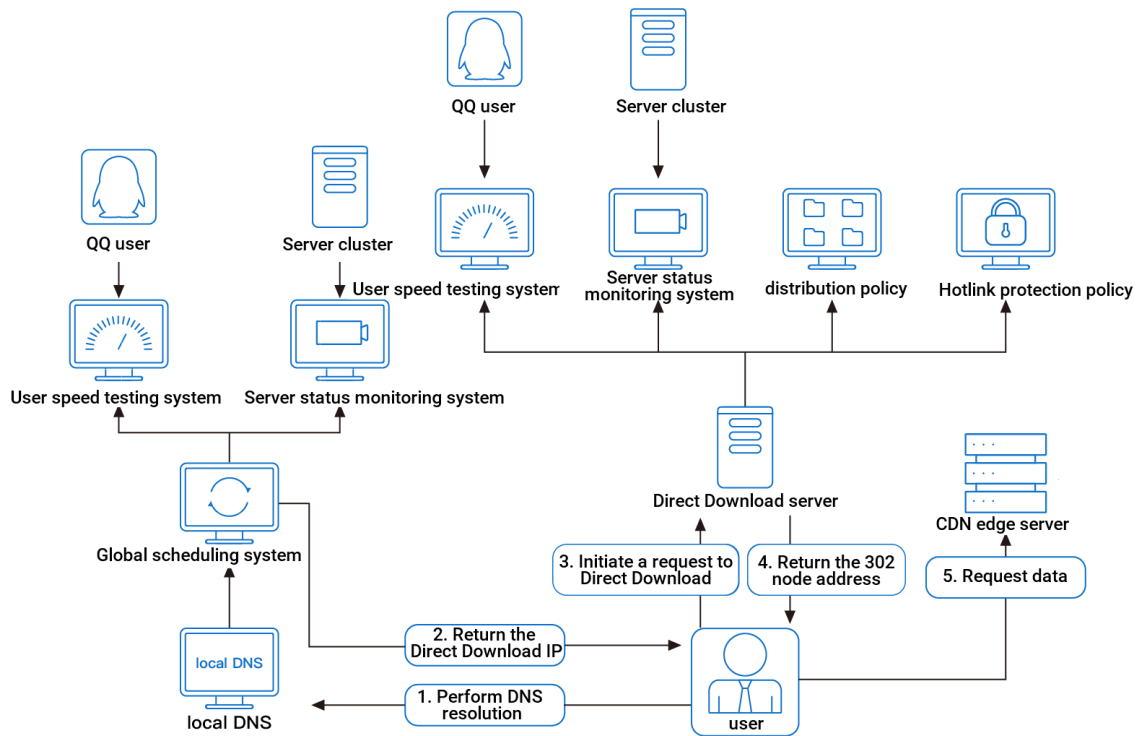


Figure 21. Dedicated DNS architecture

Dedicated DNS works as follows:

1. A user requests a URL "http://www.test.com/a.jpg". Because the domain name "test.com" is connected to Tencent Cloud CDN, it will be resolved to the global scheduling system of Tencent Cloud CDN's download acceleration platform.
2. The global load balancing system will return the IP of a dedicated DNS server to the user based on the user speed test system and the monitoring status of the dedicated DNS server.
3. The user then initiates the request to the dedicated DNS server.
4. The scheduling system of the dedicated DNS server checks the following information:
 - a) Global network map from the external user speed test system (network congestion and optimal connection are determined based on the data map of the entire QQ/WeChat network);
 - b) Status of servers of the monitoring system and node devices;
 - c) File distribution policy of the configuration system;
 - d) Hotlink protection policies of the configuration system and the custom configuration system;

It then selects the optimal edge server and returns the HTTP request to the user through 302 redirect (the IP of the returned edge server is directly added to the URL), so the user can initiate the request to the edge server.

5. The user directly initiates the request to the edge server. As the request is made over IP, it can be protected from traffic hijacking.

Dedicated DNS has the following advantages compared with the DNS scheduling system:

	DNS Scheduling	Dedicated DNS
Scheduling Accuracy	It depends on the DNS scheduling system.	The scheduling system maintains its own IP library and can accurately identify any IP range to solve cross-ISP access problems caused by incorrect Local DNS configuration, especially inaccurate scheduling with small and medium-sized ISPs.
Scheduling Granularity	It can be accurate to the domain name level.	It can be accurate to the domain name, directory, and URL levels.
Effective Time for Scheduling Rules	It depends on the TTL of the domain name.	It takes effect in real time. Traffic scheduling is precise and controllable to support businesses with traffic surges.
Applicable Scenarios	It is applicable to small static resources such as webpages and images.	It is applicable to large file downloads, such as videos and application installation packages, and supports redirection-based tracking.

Table 2. Advantages of dedicated DNS

3.3.3.2 Dynamic Content Delivery Recognition

Cache on server nodes can be built through proactive delivery or passive origin-pull.

For common small files (such as images, JS files, and CSS files), passive origin-pull can greatly simplify user operations: you only need to store your content on the origin server, and CDN will take care of everything else. When the user accesses the CDN node and the node finds no locally cached copy, it will choose the optimal network path to get the file copy from the origin server and store it.

When large files such as program updates are released, however, edge/intermediate servers may not have cached data and nodes will forward requests to origin servers for origin-pull, significantly increasing the pressure on origin servers. If the prefetch feature is used, content delivery will depend on the judgment of developers and OPS engineers, and you need to initiate the prefetch from your side or through the CDN service provider. If files are not prefetched in time, traffic problems may occur. Because all nodes will pull files from the origin server if passive origin-pull is used, it takes a long time to deliver files to CDN nodes.

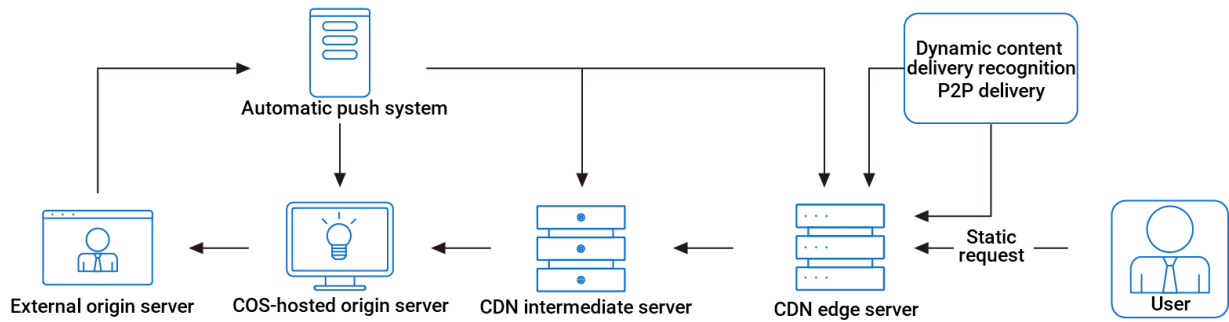


Figure 22. Dynamic content delivery recognition flowchart

The unique dynamic content delivery recognition feature of Tencent Cloud CDN can dynamically recognize traffic surges and analyze if they are caused by the delivery of large files. If yes, it will use the proactive push method to quickly push files to nodes. Tencent P2P network can accelerate the entire delivery process, significantly reducing the bandwidth consumption of origin servers.

3.4 VOD Acceleration Platform

Since 2016, major video websites have focused their strategies on creating high-quality content, while membership-based business model gains traction. The VOD acceleration platform is one of the core features of Tencent Cloud CDN's video solutions. With a complete set of key capabilities for all VOD services including upload, transcoding, management, publishing, and statistics collection, Tencent Cloud offers the one-stop Video Platform as a Service (VPaaS). Backed by flexible, fast, and high-quality video delivery service, you can focus on business development and respond to market changes with stable and reliable video delivery capabilities.

This solution has the following main advantages:

- **Experience:** it leverages Tencent's years of technical expertise and experience in the audio/video industry, which support hundreds of millions of daily playbacks in Tencent Video, millions of daily concurrent QQ audio/video calls, and more.
- **Resource:** its CDN network has over 1,300 global nodes and more than 120 Tbp bandwidth. The 10,000+ distributed transcoding clusters guarantee the transcoding quality and efficiency, which can sustain up to 2,000 concurrent transcoding tasks.
- **Channel:** it is integrated with WeChat to directly generate links for video publishing on WeChat Official Account, helping users quickly publish videos on WeChat Official Account.
- **Ecosystem:** it is connected to resources in Tencent Video V+ channel, providing users with diverse video publishing platforms and monetization methods.
- **Security:** playback speed can be controlled to build different levels of user experience. Hotlink protection based on timestamps and other solutions are supported to ensure the security and privacy of video content.

3.4.1 Applicable Scenarios

The VOD acceleration platform of Tencent Cloud CDN combines the advantages of distributed deployment with the high performance of centralized processing to address issues of high traffic consumption and long latency in streaming VOD, making it ideal for VOD streaming websites based on HTTP and HTTPS protocols, such as video websites (movies, TV shows, education, etc.), portal websites, music websites, internet TV stations, online music stations, and UGC websites.

Currently, common formats in online VOD are FLV, MP4, and HLS:

- **MP4:** it is a classic file format and well supported by mobile devices and PC browsers. It is very complicated, however, and the cost for processing is high. Due to the complexity of the index table, the loading speed for MP4 files longer than 5 minutes is slow during online playback.
- **HLS:** it is developed by Apple and works well on mobile browsers, but Internet Explorer's support for HLS depends on Flash Player. With M3U8 files formatted to index streams, it avoids MP4's drawback of slow indexing and is ideal for VOD.
- **FLV:** it is a container format developed by Adobe and commonly used by live streaming platforms. It is well supported by Flash Player on PCs, but requires a dedicated player on mobile devices. It is not supported by most mobile browsers.

3.4.2 Platform Architecture

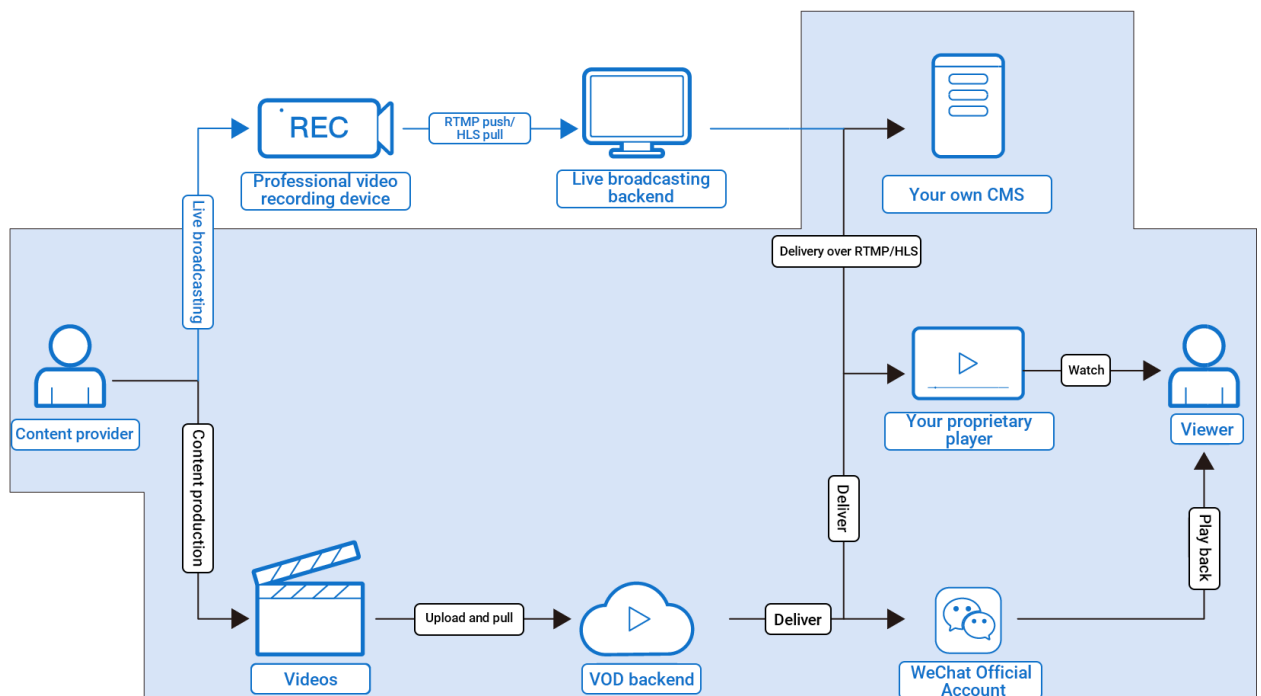


Figure 23. VOD acceleration platform architecture

As shown in Figure 23, the platform has three main components:

- **Upload component:** you can store your videos on the VOD backend by uploading or pulling

through web or SDKs and perform automatic transcoding as needed.

- **Management component:** it mainly includes the transcoding center and the console. You can manage (such as information editing, categorization and export, fuzzy search, online preview, etc.) and transcode (such as video transcoding and watermarking) video files through APIs or in the console.
- **Delivery component:** with CDN nodes deployed across China, VOD users can enjoy flexible and fast playbacks. You can use your own video player SDKs or these provided by Tencent Cloud for integration with existing businesses. You can also generate video links corresponding to dedicated WeChat Official Account for publishing.

3.4.3 Platform Features

3.4.3.1 Multichannel Upload

You can upload videos to Tencent Cloud CDN's VOD acceleration platform through different channels, including batch upload through APIs, upload through webpages, and dedicated SDK uploads for UGC. Large file upload (up to 10 GB in size) and resumable upload are supported.

Videos in the following formats can be uploaded:

Standard	Format
Microsoft	WMV, WM, ASF, ASX
REAL	RM, RMVB, RA, RAM
MPEG	MPG, MPEG, MPE, VOB, DAT
Other	MOV, 3GP, MP4, MP4V, M4V, MKV, AVI, FLV, F4V

Table 3. Supported video formats

To make it easier for you to call APIs and manage complex multipart uploads, auxiliary SDKs are provided. For more information on server-side upload SDKs provided by Tencent Cloud CDN and the integrated upload feature in the player SDKs, please see [3.4.3.3 SDK Services](#).

3.4.3.2 Transcoding Center

Tencent Cloud CDN provides transcoding services for audio/video files. It supports resolutions from 240p to 1080p and formats such as MP4 and HLS, and allows you to add custom watermarks to videos.

Transcoding audio/video files involves complex parameters. To simplify the call process, VOD acceleration platform of Tencent Cloud CDN predefines a series of standard transcoding specifications, each of which corresponds to a set of transcoding parameters. For more information on parameters, please see transcoding output specifications of the VOD system. In addition to the standard transcoding specifications, you can submit a ticket if you need custom specifications, and

the VOD team will help you configure. Table 4 shows some of the preset container formats in Tencent Cloud CDN.

Specification ID	Container Format	Video Parameters							Audio Parameters			
		Codec	Frame Rate	Width	Color Space	Bit Depth	Bitrate Mode	Bitrate	Codec	Bitrate	Channel	Sample Rate
10	MP4	H.264	24	320	YUV420P	8	ABR	256 Kbps	ACC	48 Kbps	Dual	44,100 Hz
20	MP4	H.264	24	640	YUV420p	8	ABR	512 Kbps	ACC	48 Kbps	Dual	44,100 Hz
30	MP4	H.264	24	1,280	YUV420p	8	ABR	1,024 Kbps	ACC	48 Kbps	Dual	44,100 Hz
40	MP4	H.264	24	1,920	YUV420p	8	ABR	2,500 Kbps	ACC	48 Kbps	Dual	44,100 Hz

Table 4. Examples of preset container formats

3.4.3.3 SDK Services

To facilitate video use and upload, Tencent Cloud CDN provides server-side upload SDKs. Meanwhile, to help users quickly implement VOD solution, it provides VOD player SDKs that allow users to easily build VOD sites. For user-generated content (UGC) and to keep "SecretID" and "SecretKey" from disclosure, it provides UGC-dedicated solutions and corresponding SDKs.

3.4.3.3.1 Server-side Upload SDK

Calling server-side APIs is complex. For common APIs, they require parameter sorting and signature calculation. As video files are usually large, multipart upload is used, which involves initialization, multipart upload, and completion (as shown in Figure 24).

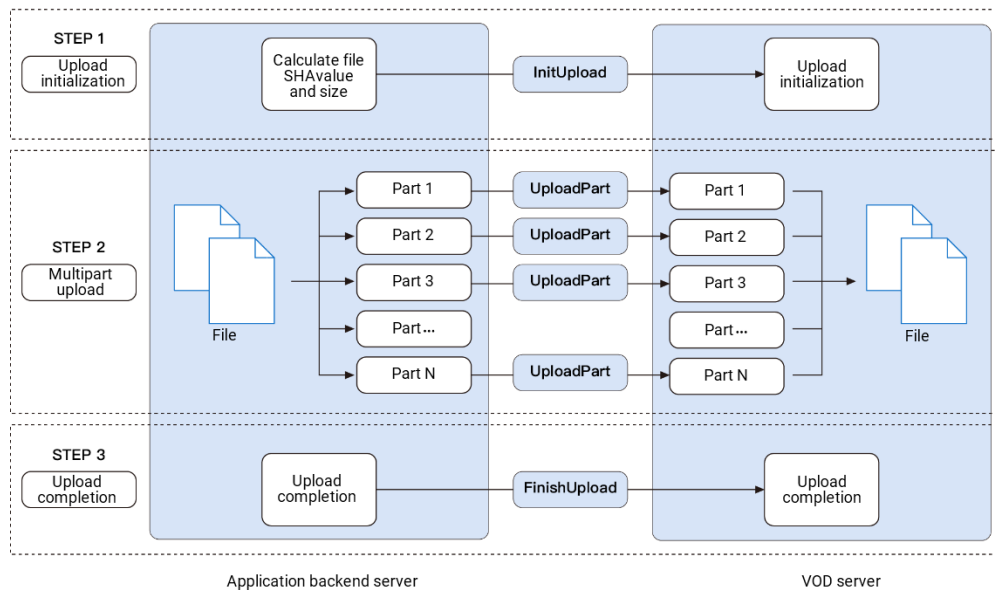


Figure 24. Steps for server-side upload

Multipart upload involves complex logics and operations. To facilitate use, VOD provides server-side SDKs to simplify the development process. The APIs have two main features: server-side APIs calls and video uploads. To keep "SecretID" and "SecretKey" from disclosure, however, for requests from clients or with signature requirements, please use the SDK as described in section [3.4.3.3.3UGC-Dedicated SDK](#).

Currently, Tencent Cloud CDN provides Tencent Cloud VOD 4.0 Server SDK for PHP so developers can use the VOD upload feature easily and quickly with their own PHP codes. Developers can specify the number of multipart to be uploaded concurrently and the number of times each part can be re-uploaded. Resumable upload is also supported. To get the latest SDKs and demos, visit <https://github.com/tencentyun/vod-php-server-sdk-v4>. SDKs for Node.js, Python, Java, Go, and C# will be available in the near future.

3.4.3.3.2 VOD Player SDK

Once a video file is stored on the server and you have the URL for playback, you can use the video player SDKs provided by Tencent Cloud CDN to play and download the video simultaneously.

In addition, Tencent Cloud SDK provides a comprehensive solution for web and mobile devices (iOS/Android versions). With flexible APIs, it can be integrated quickly with web services and applications. The content providers do not need to develop their own video players, but can directly use the built-in player in Tencent Cloud SDK to implement the VOD feature on desktops and mobile applications. The SDK also supports video upload through webs and applications.

The player provides the following features:

- Video control: play, pause, fast-forward, rewind, stop and resume a video;
- Display adjustment: video display can be adjusted to fit the entire screen or proportionally

resized;

- On-screen comment: on-screen comment can be added to the current video (no additional features are provided for on-screen comment);
- Video encryption: SDK can automatically prompt user to enter the password to play an encrypted video;
- Sharing: you can use the built-in sharing feature of the player to share a video to platforms such as Sina Weibo, or generate a QR code for video distribution;
- Adaptive bitrate streaming: video quality is adaptive to the current network condition. If a video has a certain definition, you can use the parameter to specify the definition the video should be played in.

3.4.3.3 UGC-Dedicated SDK

As the need for personalization increases among end users, simple text interaction and image upload can no longer meet demands for displaying and sharing information. As a result, UGC enters the mainstream, where video needs to be uploaded from UGC client (iOS/Android/Web) to the VOD server. Because the application should not disclose its "SecretId" and "SecretKey" to the client, upload APIs are not longer suitable. In this case, you can use UGC-dedicated APIs or SDKs (WEB/iOS/Android) for upload.

The UGC feature is implemented jointly by the SDK, Tencent Cloud COS server, and VOD service. The end user only needs to click to upload the video. For applications, upload is enabled by the SDK. You only need to integrate the SDK into your applications, and Tencent Cloud CDN will take care of the rest.

UGC video upload is a two-step process:

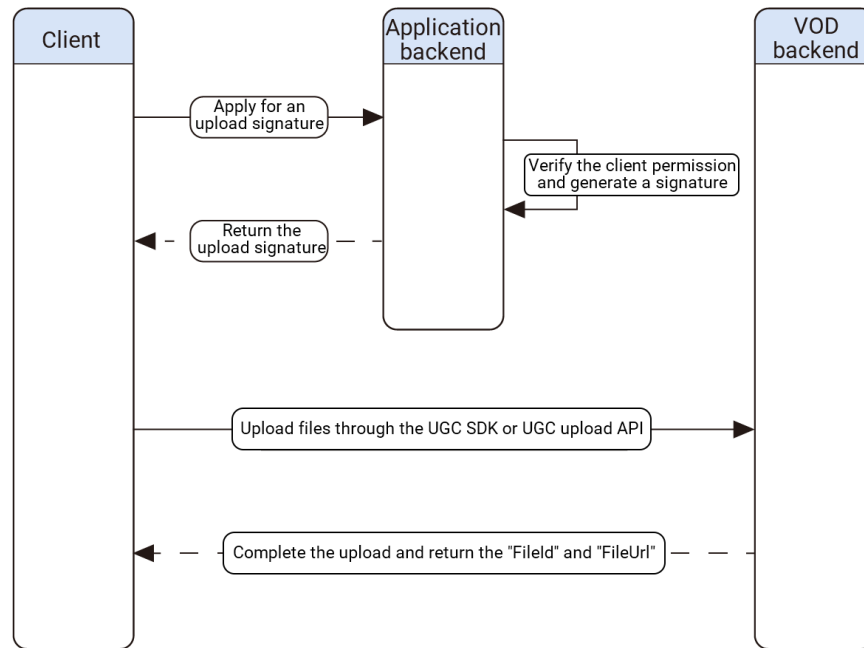


Figure 25. UGC video upload flowchart

1. Generate the upload signature: the UGC video upload signature in VOD is a Base64-encoded binary string containing the following information:
 - 1) Input parameters: includes but are not limited to the application's "SecretID", video information (video name and tags), and processing method after the video is uploaded to VOD (whether to perform transcoding, porn detection, etc.);
 - 2) HMAC-SHA1 signature generated with "SecretKey", which is used by the VOD backend to verify the validity of the UGC upload signature;
2. Based on the platform, call the UGC SDK of Tencent Cloud VOD (or call the corresponding upload API) to upload the video.

3.4.3.4 Publishing Videos on WeChat Official Accounts

Generating video links for WeChat Official Account is an important feature of Tencent Cloud CDN's VOD service. To ensure that content published on WeChat Official Account is in compliance, the platform currently only allows specified video links to be published.

With this feature, you can easily use videos already stored in Tencent Cloud CDN and generate links that can be used in WeChat Official Account messages to publish the videos. Before using this feature, make sure the Tencent Cloud account you currently operate on has been verified in "Console > Account Center > Account Information".

After identity verification, you can generate video links on the video management page. The system will automatically select an MP4 file with high bitrate for transcoding and generate a link for

it, which takes about 4 hours. After the link is generated, you can view it on the corresponding video management page under "Media Assets" and preview using a QR code. You can copy the link to the message page of your WeChat Official Account to publish the video.

3.5 Live Video Broadcasting Acceleration Platform

Given the popularity of the Internet and OTT services in recent years, demand for live video broadcasting soars. According to the 44th statistical report on China's Internet development published by CNNIC in 2019, the number of online video users reached 759 million as of June 2019, up 33.91 million from the end of 2018. Among them, 648 million were short video users. With the improvement and price cut of 4G network as well as the development of 5G, watching videos and live broadcasting content on mobile phones becomes increasingly common.

The traditional telecommunications network is a dedicated network that directly provides multicast, where the multicast protocol is used to support the entire process from the backbone network to the access network. In current Internet applications, however, OTT service providers use the public network for live broadcasting, and most of them adopt HTTP and RTMP over HTTP protocols due to firewall concerns. In this case, the live media server needs high network bandwidth to sustain millions of user requests, while the last mile connection from an ISP to the end user becomes the key in determining the live broadcasting quality. This often causes problems:

1. High latency/low definition: the competitive advantages of video business lie in playback smoothness and video definition, which are subject to the network condition. But multi-network connections across ISPs, where the access and last mile connections may be unstable, can significantly affect the user experience.
2. Traffic surge: when the content provider live broadcasts for marketing campaigns or major events and when breaking news occurs, requests for live broadcasting will often suddenly surge. The existing network may not be able to manage this.
3. Uncontrollable content: live broadcasting content is completely generated by the host and distributed to viewers in real time. In addition to ensuring that the live broadcasting is smooth and in real time, platform policies and relevant laws need to guarantee the legitimacy of the content pushed by the host. If the content violates any laws, the reputation of the platform will be damaged and it may bear regulatory consequences.

3.5.1 Applicable Scenarios

Tencent Cloud CDN's live video broadcasting acceleration platform can help you balance latency and video definition, respond to traffic surges, monitor content, etc. It improves the live broadcasting quality and user retention, while ensuring the security and content compliance of the platform.

Live broadcasting acceleration is suitable for streaming media applications and webpages, such as live game broadcasting, live shows, news, sports events, online education, video conferences, etc.

3.5.2 Platform Architecture

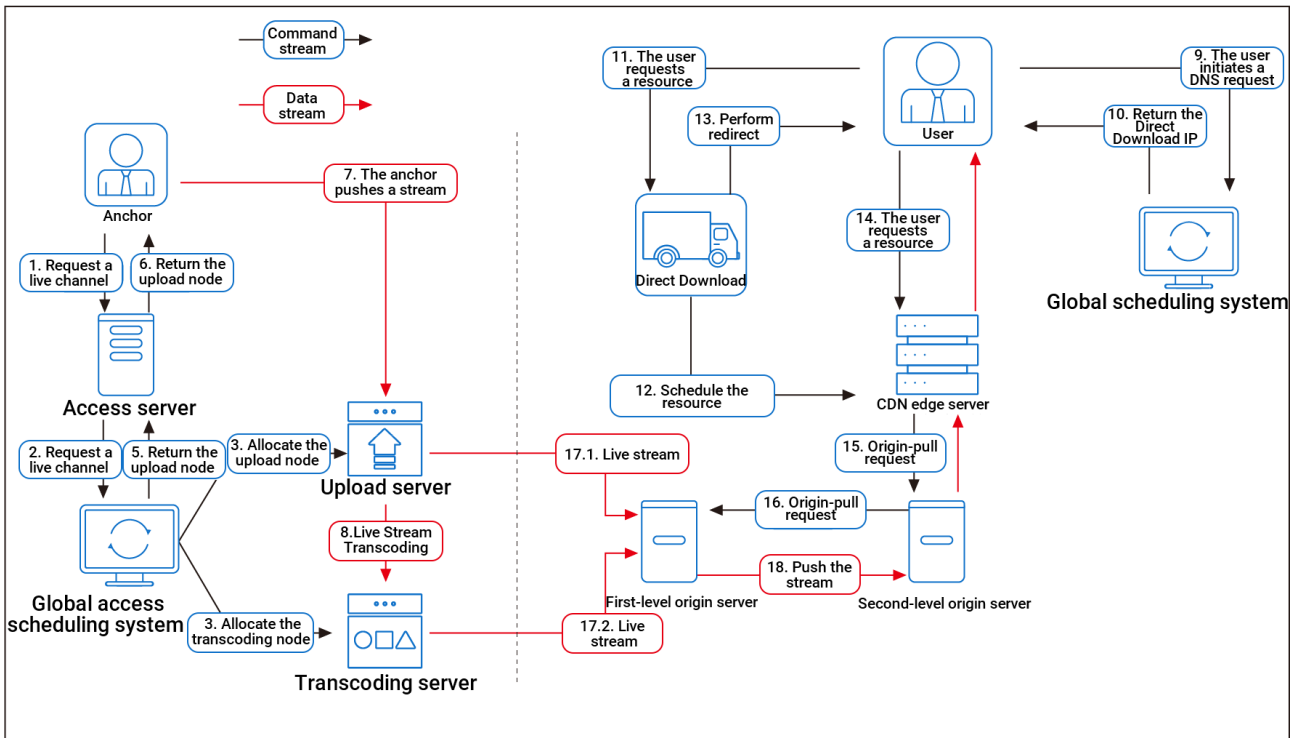


Figure 26. Live video broadcasting acceleration platform flowchart

Tencent Cloud CDN's live broadcasting acceleration platform involves two parts: host access and viewer access, as shown in Figure 26. The following describes each process in detail:

1–6: an access request from the host will not be directly transferred to the upload server. Instead, the host needs to connect to the "access server" first, which will work with the global access scheduling system to select the optimal upload and transcoding servers based on the geographical location of the host and the current server load/speed. The information will be returned to the host;

7–8: the host client will push the video to the upload server, which will transfer the live streams to the transcoding server at the same time. Based on the live broadcasting protocol and business requirements, it will divide and transcode the bitstream into .ts segments under the HLS protocol, or get the HLS .m3u8 playlist by the live stream and pull the HLS .ts segments accordingly. The live stream can be transcoded to multiple bitstreams for output. Tencent Cloud transcoding system not only supports transcoding live streams, but also encoding videos uploaded to VOD. Based on business needs and customer requirements, you can select custom bitrates and provide your customers with personalized multi-bitrate services. Parameters such as bitrate, output format, and resolution can be customized;

9–14: similar to the process as described in section [3.3.3.1 Dedicated DNS](#), when an end user accesses from the DNS system, dedicated DNS node rather than an edge server will be allocated to the user by the global scheduling system. Dedicated DNS will forward the user request to an edge server to request the data;

15–16: if the edge server has no cached data, it will initiate a request to upper-level nodes for

the data;

17–18: output data streams from the upload server and transcoding server will be forwarded to the end user from top down and cached on nodes.

3.5.3 Platform Features

3.5.3.1 Multichannel Standby Mechanism

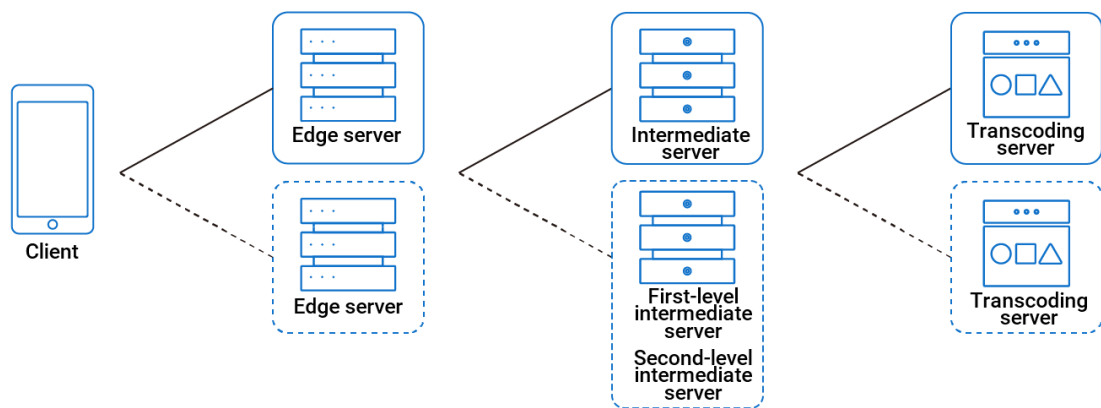


Figure 27. Standby mechanism

Tencent Cloud CDN offers a standby mechanism to ensure smooth live broadcasting:

1. On an edge server, if an exception such as server failure occurs or maintenance operation such as version upgrade and routine OPS is performed, client will request a standby edge server based on the missing data segments and splice the obtained data to ensure the continuity of video streams.
2. If an intermediate server fails, it will switch to another intermediate server (level-1 or level-2) based on the private protocol or .ts segment number to achieve seamless switchover.
3. If a transcoding server fails, the intermediate server will switch to a standby transcoding server, achieving seamless master/slave switchover for live broadcasting. Tencent Cloud CDN configures master and slave transcoding servers in the module. The master server continuously encodes and outputs streams, while the slave server pulls the source data without encoding it. When the master server fails, the upload server will switch to the slave server and pass in the serial number of the currently required frame. Then, the slave server will start transcoding from this frame on (because the original data has already been cached) and take over the transcoding task.

3.5.3.2 Elastic System

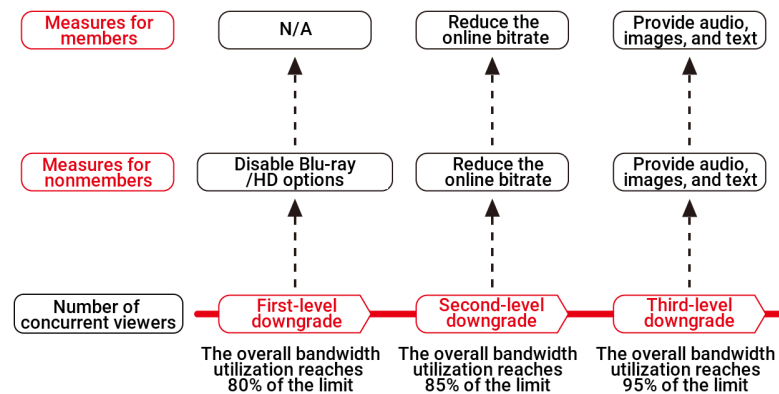


Figure 28. Multi-level elastic system

As shown in the figure above, in addition to ensuring quality and supporting as many users as possible, Tencent Cloud CDN's live broadcasting acceleration solution has an elastic system on standby. When traffic surges and bandwidth utilization rises sharply, the elastic system will help ensure a smooth viewing experience before the resource pool is modified. In this scenario, the following policies can be customized:

1. When the overall bandwidth utilization of a live broadcast reaches 80%, it will have a lower definition to slow down bandwidth consumption, while nonmembers and members will have different services. For nonmembers, SD playback will be provided but Blu-ray and HD options are no longer available. Members can still watch videos in all definitions.
2. When the overall bandwidth utilization of a live broadcast reaches 85%, the elastic system will lower bitrates for all by 20%. After the reduction, some bandwidth resources can be released to support access requests from new users.
3. When the overall bandwidth utilization of a live broadcast reaches 95%, video playback is no longer available. Only audio, image and text content will be provided. New users can no longer watch videos, but they can still listen to audios and view images.

3.5.3.3 Instant Broadcasting

Interactive live video broadcasting requires instant broadcasting within 1 second, so viewers can browse hosts and switch between live rooms instantly. Tencent Cloud CDN provides the following optimizations for instant broadcasting:

1. Edge server acceleration: when the user enters a new live room, the request will be scheduled to the node closest to the user and transferred over TCP protocol to speed up data push to the client.
2. Keyframe cache: when the user enters a new live room, the latest cached key frame will be provided, which is the I frame in the figure below. After the client gets the key frame,

playback can be started. By limiting its cache size, the client can be optimized to meet playback requirements.

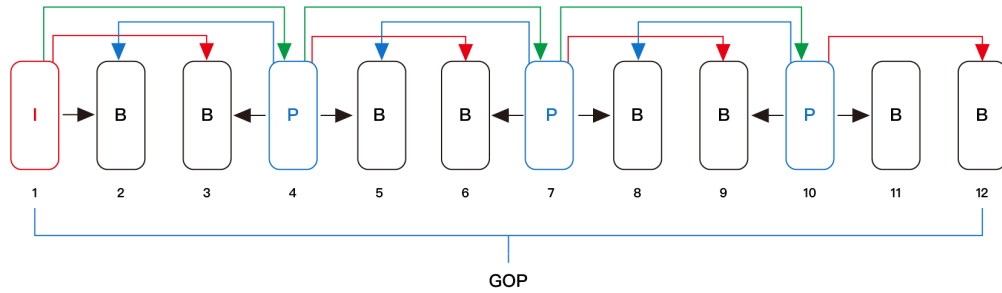


Figure 29. Live broadcasting key frame

3. Under HLS protocol, three .ts segments will be returned when a user enters a new live room, and only with all three segments can the player start playback. This is a unique characteristic of iOS player. If only one .ts segment is provided in .m3u8 file, the client will request the .m3u8 file for new .ts segments. If the .m3u8 file is not yet updated and .ts segments are original ones, the player will be triggered to wait for a timeout period before requesting an update again, resulting in slow start.

4 Conclusion

With the rapid development of the Internet, people have increasingly high requirements for information services, while looking for faster response times and better product experiences.

Backed by solid experiences in the Internet industry, Tencent Cloud CDN has gained technical expertise and strong capabilities over the years. With its vast customer base, Tencent Cloud provide services with fast response, seamless user experience, and high security to massive amount of users. Tencent Cloud CDN can help you speed up website access, alleviate pressure on servers and bandwidth resources, enhance website security, and ensure business continuity.

To help provide your end users with optimal experiences and improve user satisfaction, Tencent Cloud CDN will be your ideal choice.

5 Relevant Links

[Tencent Cloud CDN Product Page](#)

[Tencent Cloud CDN Console](#)

[Tencent Cloud CDN Product Documentation](#)

[Tencent Cloud CDN API Documentation](#)