SANGFOR

Sangfor Application Delivery Controller Technical White Paper

Sangfor Technologies Inc. 2017

Declaration

Disclaimer

This document was prepared by Sangfor Technologies Inc. The information obtained herein is provided on an 'as available' basis. Sangfor may make improvement or changes in this document, at any time or without notice.

The information is believed to be accurate. However, Sangfor shall not assume responsibility or held liable for any loss or damage resulting from omissions, inaccuracies or errors contained herein.

Contact Us

For any feedback or suggestion, please contact us through the following:

Address: Block A1,Nanshan iPark,No.1001 Xueyuan Road,Nanshan District,
Shenzhen,Guangdong Province,P.R.China(518055).
Hotline: +60 12711 7129 (7511)
Email: sales@sangfor.com/marketing@sangfor.com
Latest technology and product information is available at the official website of SANGFOR: www.sangfor.com.

Document Conventions

Abbreviation	Full Name
ACL	Access Control List
ADC	Application Delivery Controller
BRAS	Broadband Remote Access Server
DNAT	Destination NAT
DNS	Domain Name Service
DR	Direct Route
FTP	File Transfer Protocol
HA	High Availability
HTTP	Hypertext Transfer Protocol
ICMP	Internet Control Message Protocol
ISP	Internet Service Provider
MAC	Media Access Control
NAT	Network Address Translation
OSPF	Open Shortest Path First
RADIUS	Remote Authentication Dial In User Service
RIP	Routing Information Protocol
RTT	Round Trip Time
STP	Spanning Tree Protocol
SNAT	Source NAT
SNMP	Simple Network Management Protocol
SOA	Service Oriented Architecture
SSL	Secure Socket Layer
TCP	Transmission Control Protocol
UDP	User Datagram Protocol
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
VLAN	Virtual Local Area Network

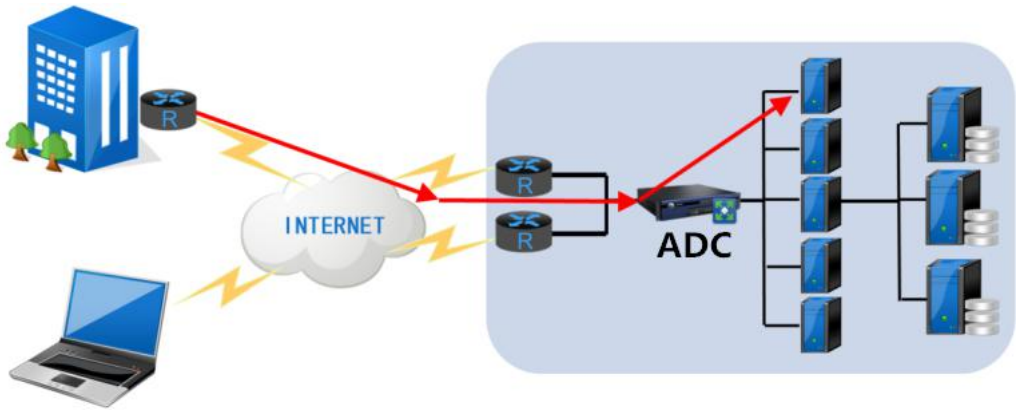
Contents

1. Background

Traditional load balance technologies help enterprise user improve data processing stability by increasing bandwidth and throughput of network devices and servers, yet with less intelligent and optimization functionality that are addressed by application delivery controller(ADC). Sangfor ADC come to cope with deployment and service delivery challenges in complex environment. By properly deploy the application delivery controller, enterprise customers can improve availability, performance and security of their business systems and applications, and make data center infrastructure even more efficient to catch up with the trends in development of software-defined data center.

1.1 Load Balance's Concentration

Initially, server load balance comes out with the purpose of handling network issues, for example, distributing requests to a server out of a group of servers that are in charge of Web application delivery. At first, round-robin DNS is used, but it has some limitations. Therefore, load balances with specific functions spring up, which can analyze inbound requests and map those requests to available servers dynamically.



In order to meet increasing complex requirements, load balance technologies have gone through a bunch of innovations, from inbound issues, such as identify workload and failures of servers dynamically, to the efforts of keep session persistence work to store user data throughout the life of a session. However, the market rapidly changed and attention has been paid to other aspects such as efficiency of applications and servers. The best practice used in this aspect is SSL offloading.
Then, focus shifts to outgoing traffic. A great many technologies and functions mushroomed, aiming at improve efficiency in application and service delivery across the Wide Area Network(WAN).
Innovation also shifts from network technologies aiming at improve efficiency of infrastructure to performance optimization and security of applications. Meanwhile, load balance starts to develop into application delivery which covers comprehensive aspects such as network, servers, applications and even security.

1.2 Advent of Application Delivery

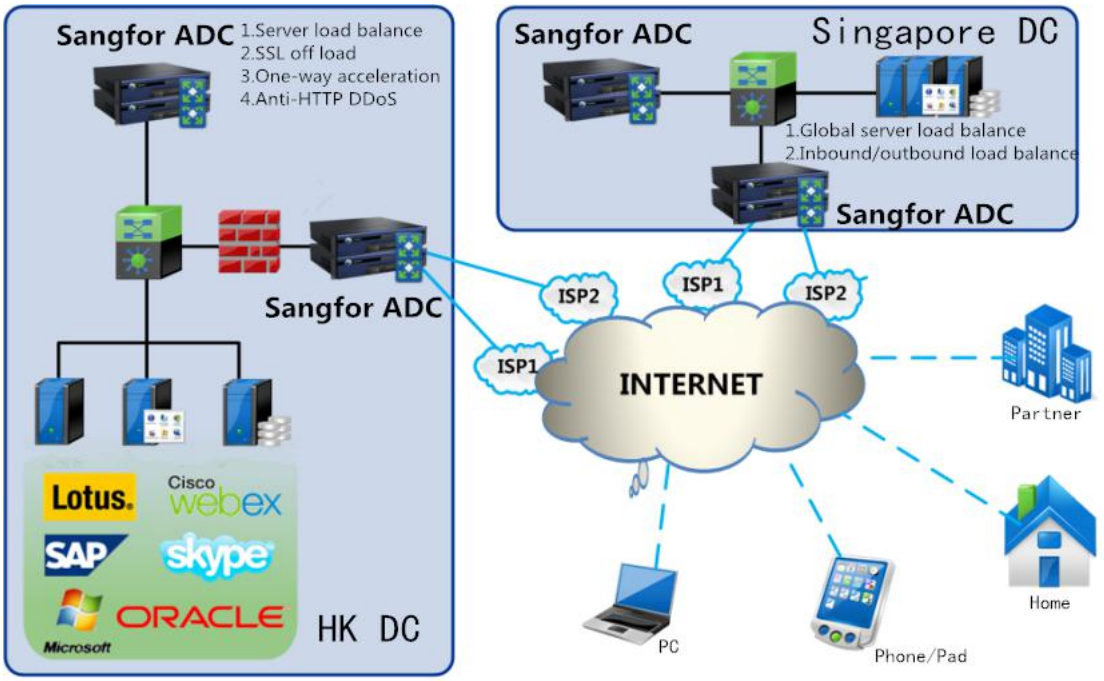
Advanced application delivery controller(ADC) can cope with the challenges brought about by complex deployment environment and application delivery. The past decade has witnessed the change of enterprise application to web-based to simplify business processes and improve productivity, as well as the development of SOA(Service-Oriented Architecture), Web2.0 and cloud computing.
More breakthroughs are made in ADC to improve application environment. In terms of business sustainability, user experience, data center availability and so on, ADC also made optimization of them possible. Last but not least, ADC greatly reduces the number of servers, provides real-time control mechanism to support data center virtualization and lowers requirements of data center for power supply and cooling, making businesses more compact, energy-saving and environmental-friendly.

1.3 Fast and Intelligent Application Delivery Network

Until now, application delivery technologies are mature. There is no significant differences among products by different vendors. The difference is focuses, some provide a wholesome solution, some address integration and some focus on security. As for Sangfor, we are to provide a fast and intelligent application delivery solutions that are specific for customers from different regions, to enable customers to gain profits more than those brought about by any other vendors, to improve user experience, and to enhancing competitiveness of their own business system as well.

2. Sangfor Application Delivery Solution

Sangfor Application Delivery Controller(ADC) provides customer a comprehensive solution, include load balance among multiple data centers, links and servers. Together with such functions as functionality optimization, one-way acceleration and multiple intelligent management features, etc, Sangfor ADC can real-time monitor status of those data centers, links and servers, and schedule client requests to corresponding data centers, links and servers based on pre-defined rule to have data flow split more properly and that all the data centers, links and servers can be made full use of. Moreover, it also helps to enhance overall capability and stability of the application system, improve user experience and reduce IT investments.



2.1 Concepts

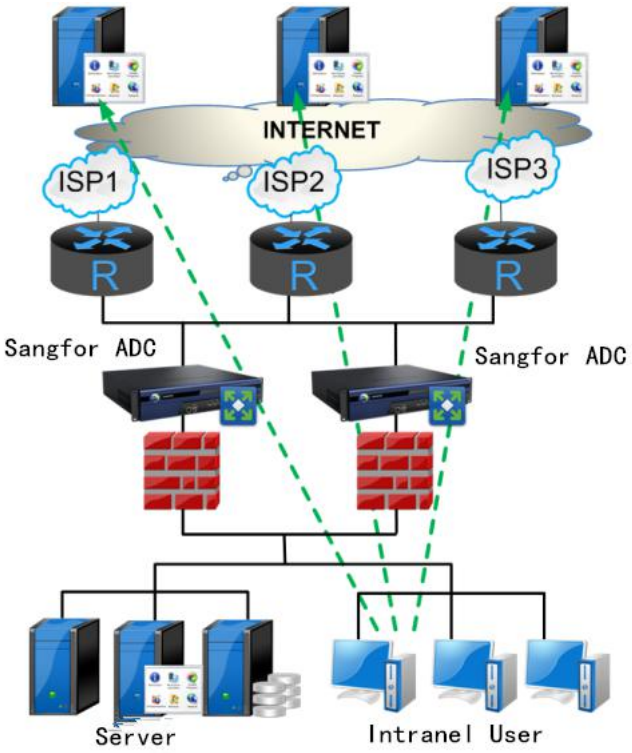
- Load balance Algorithms: The load balance algorithm defines the criteria that the Sangfor ADC used to select the service to which each client requests are scheduled, including select of LB-enabled servers, links and site.
- Health Check: This is a practice of probing health (availability) of servers or links.
- Proximity: In link load balance scenarios, Sangfor ADC will select the optimal link based on such factors as client location, ISP and health status of links, etc, so as to ensure that traffic goes through the optimal link.
- Virtual Service: Service delivered by Sangfor ADC is virtual service. As for virtual service, such properties as service type(protocol), node pool(IP addresses and ports of one or multiple servers), etc., should be configured. When a client request reaches Sangfor ADC, it may match a virtual service on Sangfor ADC and then be directed to a real server based on a load balance policy.
- Session Persistence: This can ensure that client requests are directed to the same server in a pool throughout the life of a session or during subsequent sessions, by identifying, tracking and store specific type of information. In that case, the predefined load balance method will not be applied.
- IP Address Database: A database contains IP ranges of different Internet Service Providers(ISPs), with which, ISP information can be found based on the source IP address or destination IP address of messages, then an optimal link will be selected for the traffic.

2.2 Multiple Links Load balance

Sangfor ADC integrates such algorithms as inbound load balance, outbound load balance, intelligent DNS resolution, Round Robin, Weighted Round Robin, static proximity, dynamic proximity, etc, which helps to solve the problem of traffic control over multiple links, enhances bandwidth usage and helps enterprises to reduce investment in links. Meanwhile, it offers best user experience by selecting an optimal link. Moreover, Sangfor ADC use such technologies as link health check and session persistence and makes full use of the high reliability provided by multiple links, so that service will not be disrupted even when certain link is down and optimal user experience can be provided.

2.2.1 Outbound Load balance

When LAN users request for external resources, Sangfor ADC will distribute traffic to different WAN links based on the pre-defined load balance policies after receiving requests from LAN users, so as to enhance bandwidth usage of WAN link.

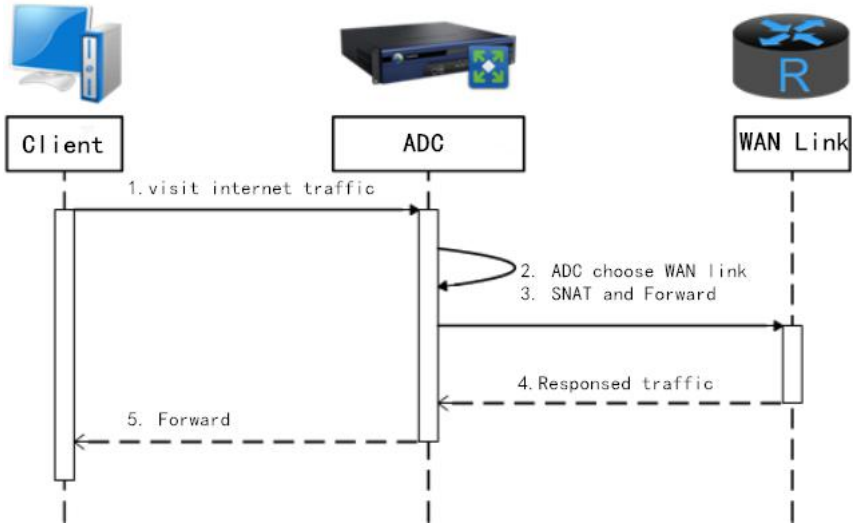


1. Realization

After receiving requests from LAN users, Sangfor ADC will selects optimal outbound links based on pre-defined load balance policies(Bandwidth,bandwidth usage,link status,source/destination IP,domain,etc...) and translate source IP address(Translate one valid source IP address or do automatic mapping with an interface IP address of Sangfor ADC), so as to ensure that returned packets can be received successfully.

2. Work Flow

Work flow of outbound load balance:

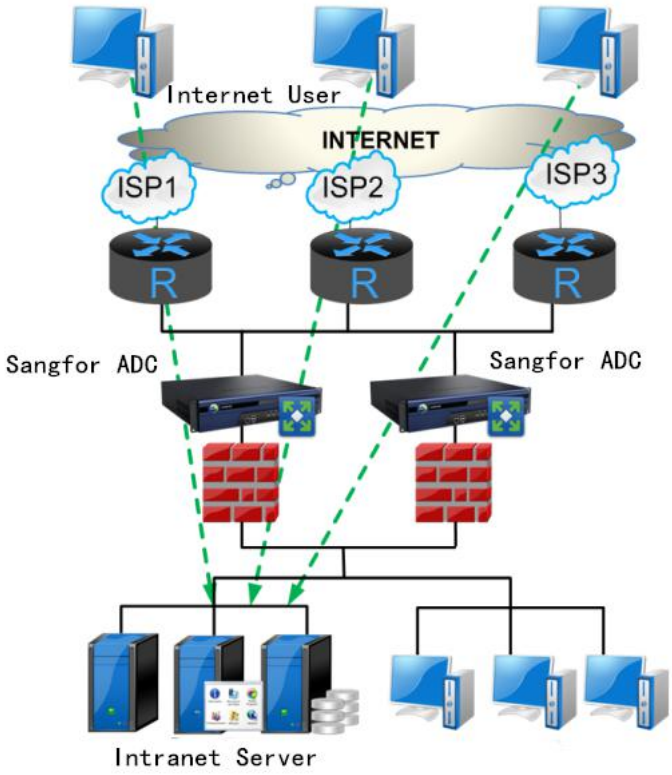


Detailed description of the work flow:

Steps	Description
1	Sangfor ADC receives requests from LAN users.
2	Sangfor ADC selects optimal outbound links based on pre-defined load balance policies.
3	Sangfor ADC assigns traffic to the outbound links that have been selected and translates source IP addresses.
4	Sangfor ADC receives traffic returned from external network.
5	Sangfor ADC forwards traffic to LAN users.

2.2.2 Inbound Load balance

When WAN users request for internal resources, Sangfor ADC uses the technology of intelligent DNS, supports assigning multiple public IP addresses of different ISP for one domain, and resolve different ISP user DNS requests. Based on load balance policies, Sangfor ADC returns optimal address to clients of different ISP user and ensures that bandwidth is shared and distributed in a responsible and intelligent way.



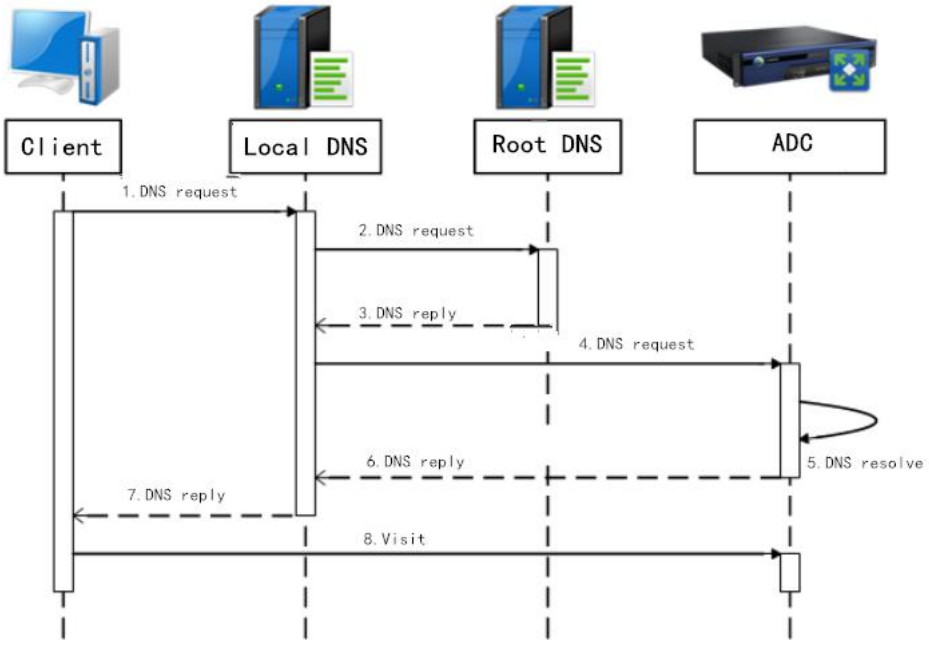
1. Realization

DNS rights should be obtained by going to the domain provider to change NS records. Sangfor ADC supports assigning multiple public IP addresses of different ISP for one domain, and resolve different ISP user DNS requests.

With different load balance policies provided by Sangfor ADC, users can access to internal resources through different links, Sangfor ADC supports reverse query with multiple links and can check link status based on Round-trip time(RTT), and then return best IP address based on the two factors.

2. Work Flow

Work flow of inbound load balance:



Detailed description of the work flow:

Steps	Description
1	Client of WAN user sends DNS request to local DNS server.
2	Local DNS server searches locally whether there are relevant records. If not, it will send queries to root DNS server.
3	Root DNS server responds to local DNS server and tells that DNS rights have been delegated to Sangfor ADC.
4	Local DNS server sends DNS requests to Sangfor ADC.
5	Sangfor ADC checks link status first, and then select an appropriate IP address based on the pre-defined load balance algorithm.
6	Sangfor ADC sends the IP address to local DNS server.
7	Local DNS server forwards the IP address to client.
8	Client sends request to LAN server with the IP address.

2.2.3 Load balance Algorithms

Sangfor ADC provides the following algorithms for administrators to select based on different needs.

1. **Round Robin**
   Mechanism: All links are in a queue, Sangfor ADC choose link for session in a periodically repeated order.
   Scenario: The links are from the same ISP, and the bandwidth of the links is similar.
2. **Weighted Round Robin**
   Mechanism: Based on throughput of links, each link is assigned a weight, an integer value that indicates the processing capacity. Links with higher weights receive new connections first than those with less weights, and links with higher weights get more connections than those with less weights and links with equal weights get equal connections.
   Scenario: The links are from the same ISP, and the bandwidth of the links is not the same.
3. **Weighted Least Connection**
   Mechanism: In the weighted least-connections scheduling, new network connection is assigned to a link which has the least ratio of the current active connection number to its weight.
   Scenario: Bandwidth of the links is not the same and the connections initiated by different users are kept for a period time which is quite different from each other.
4. **Weighted Least Traffic**
   Mechanism: In the weighted least-traffic scheduling, new network connection is assigned to a link which has the least ratio of the real-time traffic to its weight.
   Scenario: There are multiple links and the bandwidth of the links is quite different from each other.
5. **Static Proximity**

Mechanism: Link is selected based on the pre-defined static proximity rule, or is selected based on the built-in global IP address database.

Scenario: The links are from different ISP and traffic is mainly inbound traffic.

6. **Dynamic Proximity**

   Mechanism: An optimal link will be selected based on delay in data transmission and real-time load of the link.

   Scenario: The links are from different ISP and traffic is mainly outbound traffic.

7. **Bandwidth Ratio**

   Mechanism: Since the throughput of different links are different, each link is assigned with a weight based on bandwidth, and client requests are directed to links based on the weight.

   Scenario: The links are from the same ISP, but the bandwidth of the links is quite different from each other.

8. **Hashing**

   Mechanism: Requests of different clients will be distributed to different links based on hash value of the IP address of local DNS server.

   Scenario: There are multiple links, and requests of the same client should be distributed to the same link.

9. **Primary/Secondary**

   Mechanism: There are primary and secondary links, and requests will be scheduled to the secondary link when the primary link fails.

   Scenario: There are multiple links, and there is a high requirement for service continuity.

10. **First Available**

    Mechanism: Network connection will be assigned to the first valid link.

    Scenario: There are multiple links, and there is a high requirement for response speed.

2.2.4 Link Health Check

Sangfor ADC supports checking link status by checking accessibility to multiple websites. For example, check whether www.google.com, www.facebook.com are reachable using ISP1, and execute OR operation as the results. Therefore, as long as one of those website is reachable, the link is working properly. This method avoids the limitations of checking using ICMP and also avoids mistakes that may be brought about by checking a single website.

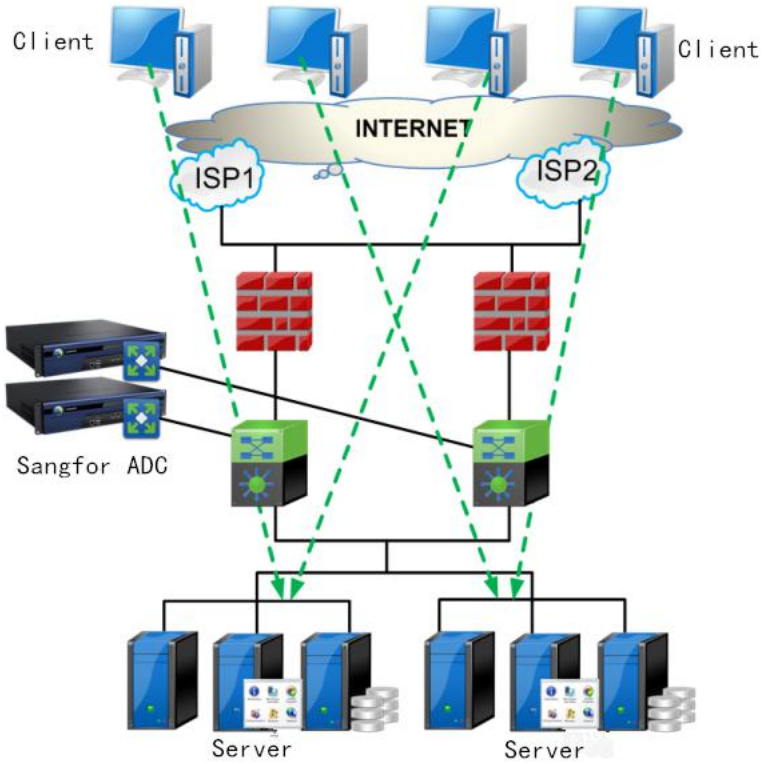
2.3 Server Load balance

Based on server cluster mechanism, Sangfor ADC supports load balance by providing virtual services based on real servers, and provides virtual IP addresses for clients. According to pre-defined scheduling policies based on multiple layer 4 and layer 7 load balance algorithms, Sangfor ADC supports distributing requests to a particular server when those requests reach ADC, so as to optimize resource use. Meanwhile, Sangfor ADC helps to solve the performance bottleneck of a single server with

reduced hardware investment, paves the way for expansion in the future, and guarantees performance in case of large concurrent sessions.

Sangfor ADC supports real-time monitor servers, and requests will be distributed to another server when server failure, so as to realize redundancy and ensure stability and reliability of crucial application systems and also service continuity in case of server failure.

Due to IPv4 address exhaustion and the increasing security and reliability in IPv6 address, IPv6 is more and more widely used in recent years. Therefore, Sangfor ADC not only supports load balance for application systems based on IPv4, but also supports IPv6-based layer 4 and layer 7 server load balance.



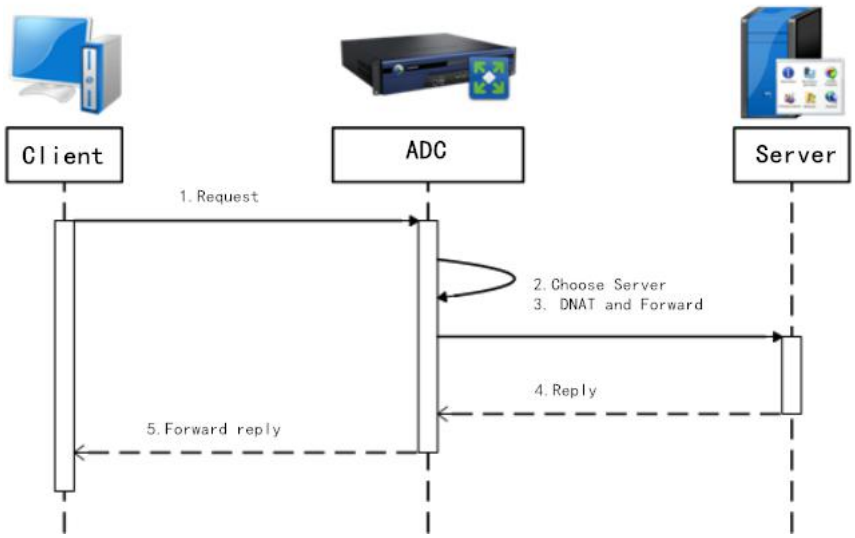
2.3.1 Layer 4 Load balance Using DNAT

1. Realization

Sangfor ADC supports load balance based on such aspects as IP address, application type, application contents, etc. Traffic of different types of applications will be distributed to different servers. Applications based on such protocols as TCP, UDP, IP, DNS, E-mail, FTP, HTTP, RADIUS, etc are supported. In layer 4 load balance, client requests are forwarded to servers by Sangfor ADC, and then TCP connection will be established between client side and server side and under this circumstance Sangfor ADC functions as a router. As for loading balance using DNAT, Sangfor ADC will translate destination IP address first and then forward client requests to back-end servers.

2. Work Flow

Work flow of layer 4 load balance using DNAT:



Detailed description of the work flow:

Steps	Description
1	Client sends requests. Source IP is client IP address and destination IP address is virtual service IP address.
2	After receiving requests, Sangfor ADC will use the pre-defined load balance scheduling algorithm to select a particular server to which to redirect each client request.
3	Sangfor ADC distributes requests using DNAT. Source IP is client IP address and destination IP address is server IP address.
4	Server receives requests and responds to requests. Source IP is server IP address and destination IP is client IP address.
5	Sangfor ADC receives responses from server and forwards responses to client side after translating source IP address. Here source IP is virtual service IP and destination IP is client IP.

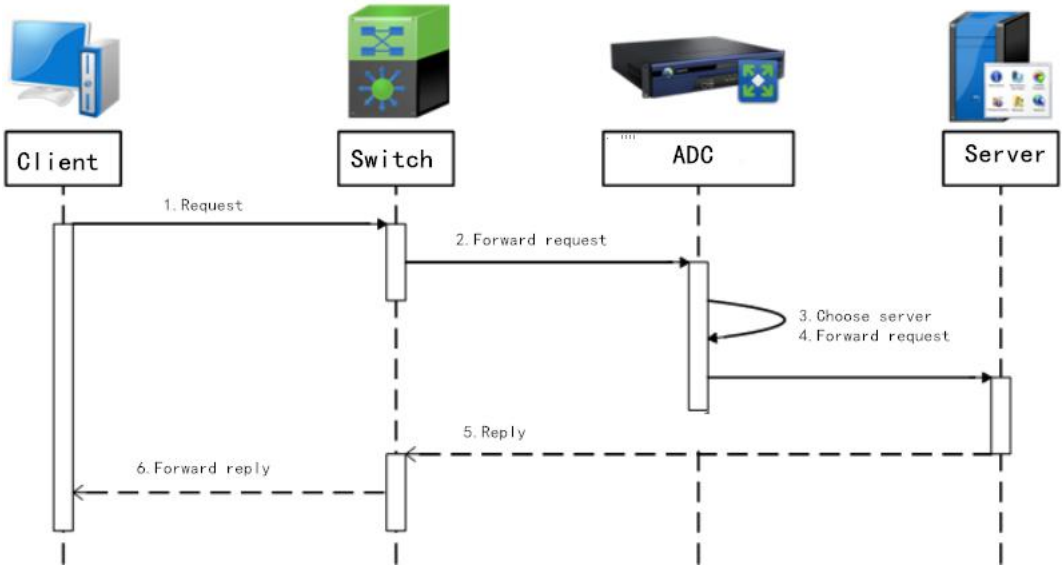
2.3.2 Layer 4 Load balance in Direct Routing(DR) Mode

1. Realization

As for load balance in Direct Routing(DR) mode, virtual service IP address should be configured on the loopback interface of back-end server. When Sangfor ADC distributes client requests, it will not change destination IP address but change destination MAC address to server MAC address, and then it forwards requests to back-end servers. Responses from back-end servers will then be directly returned to clients through switch not through Sangfor ADC. Therefore, pressure upon ADC will be reduced and performance bottleneck of the overall business system can be avoided.

2. Work Flow

Work flow of layer 4 load balance in DR mode:



Detailed description of the work flow:

Steps	Description
1	Client sends requests. Source IP is client IP address and destination IP address is virtual service IP address.
2	Switch forwards client requests to Sangfor ADC.
3	After receiving requests, Sangfor ADC will use the pre-defined load balance scheduling algorithm to select a particular server to which to redirect each client request.
4	Sangfor ADC distributes requests to servers. Source IP is client IP address, destination IP address is virtual service IP address, and destination MAC address is server MAC address.
5	Server receives requests and responds to requests. Source IP is virtual service IP address and destination IP is client IP address.
6	Switch receives responses and forwards to client side directly.

2.3.3 Layer 7 Server Load balance

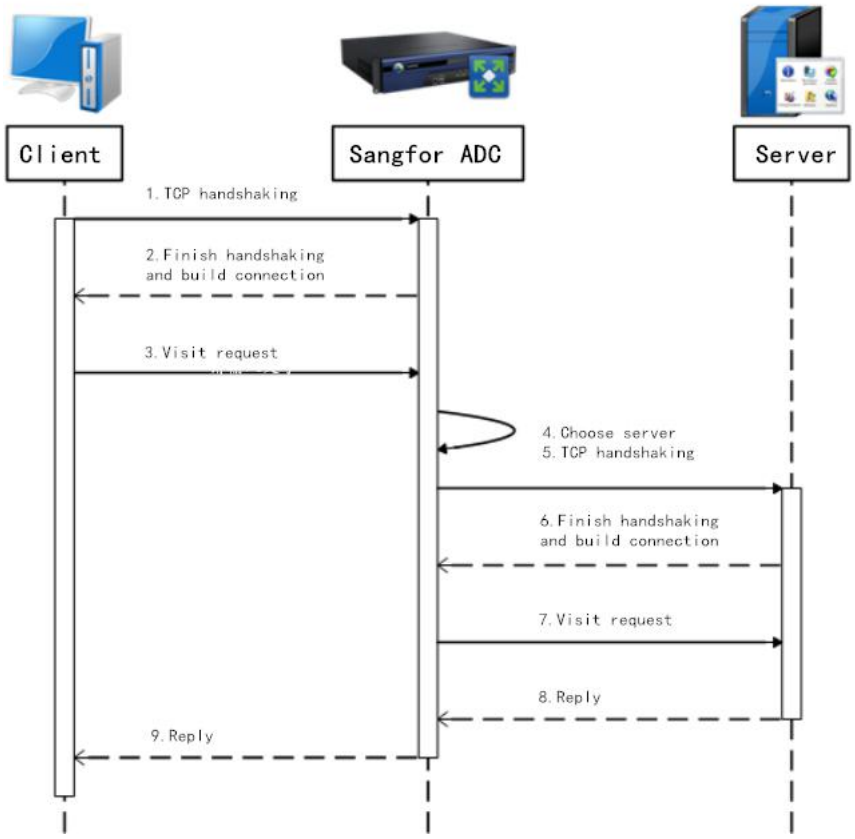
1. Realization

Administrators allocate resources based on the content exchange mechanism on the application layer, so as to cope with the complex and personalized scheduling requests. For example, server is selected based on policies such as URI, HOST, COOKIE, USER_AGENT, etc, or redirect page or discard packets by rewriting HTTP request header or response header, so as to realize interaction and association among different business systems. In layer 7 server load balance, Sangfor ADC establishes

TCP connection with client to obtain request messages first, then it selects a particular server based on application layer contents contained in the messages and establishes TCP connection with the server. Here Sangfor ADC functions as a proxy server.

2. Work Flow

Work flow of layer 7 server load balance:



Detailed description of the work flow:

Steps	Description
1	Client sends TCP requests to Sangfor ADC. Source IP is client IP address and destination IP address is virtual service IP address.
2	Client and Sangfor ADC establishes TCP connection.
3	Client sends requests. Source IP is client IP address and destination IP address is virtual service IP address.
4	After receiving requests, Sangfor ADC will use the pre-defined load balance scheduling algorithm to select a particular server to which to redirect each client request. Meanwhile, request data will be cached.
5	Sangfor ADC sends TCP requests to server. SYN seq is client SYN seq, source IP is client IP address, and destination IP address is server IP address.
6	Sangfor ADC and server establishes TCP connection.
7	Sangfor ADC changes destination IP address and TCP sequence number of the cached request data, and sends requests to server.

8	Server receives requests and responds to requests. Source IP is server IP address and destination IP is client IP address.
9	Sangfor ADC changes source IP address and TCP sequence number of responses from server, and forwards responses to client side. Here source IP is virtual service IP and destination IP is client IP.

2.3.4 Server Load balance Algorithm

Sangfor ADC uses multiple load balance algorithms to distribute traffic to servers evenly, so as to make best use of server resources and avoid load imbalance among different servers.

1. **Round Robin**
   Mechanism: Client requests are distributed to each server in a periodically repeated order based on request sequence. When server failure, requests will not be distributed to the server until the server returns to normal.
   Scenario: Performance of the servers in the cluster is similar to each other.
2. **Weighted Round Robin**
   Mechanism: Performance of servers in the cluster is different from each other, therefore, servers are assigned with different weights. Then client requests will be distributed to servers based on weights of the servers.
   Scenario: Performance of the servers in the cluster is quite different from each other.
3. **Weighted Least Connection**
   Mechanism: In the weighted least-connections scheduling, new network connection is assigned to a server which has the least ratio of the current active connection number to its weight.
   Scenario: Performance of the servers is not the same and the connections initiated by different users are kept for a period time which is quite different from each other.
4. **Fast Response**
   Mechanism: Reassign weight to servers based on response time. Server with less response time is assigned with large weight, and server with more response time is assigned with small weight. Therefore, more network connection requests will be assigned to servers with less response time, yet some network connection requests will also be assigned to servers with more response time so as to avoid load imbalance.
   Scenario: Topology of server cluster is distributed and clients will select the nearest server.
5. **Dynamic Feedback**
   Mechanism: Set standards for load balance based on such factors as CPU, I/O, memory, etc., of servers so as to dynamically change weight of each server, and select an optimal server for new network connection.
   Scenario: Processing capability of each server in the cluster are different, moreover, it is difficult to settle down weight of the servers.
6. **Hashing**
   Mechanism: Hash algorithm, based on URI, HOST, SRC_IP and IP＋PORT, helps to evenly distribute requests which contain different elements to different servers in the cluster.
   Scenario: Requests which contain the same elements should be distributed to the same server.

7. **Priority**

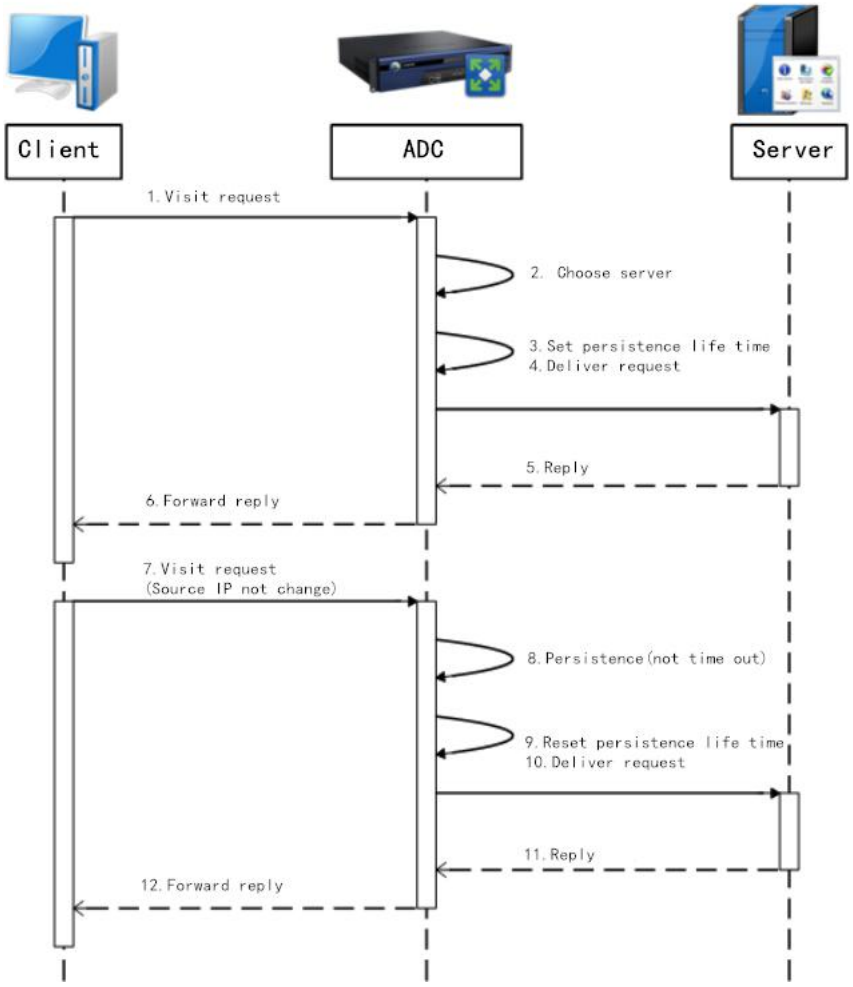
Mechanism: Servers are allocated into different groups based on priority. Requests will be distributed to servers with high priority first, and will be distributed to servers with low priority only when servers with high priority fail.

Scenario: Such factors as performance, stability, etc., of the servers in the cluster are different.

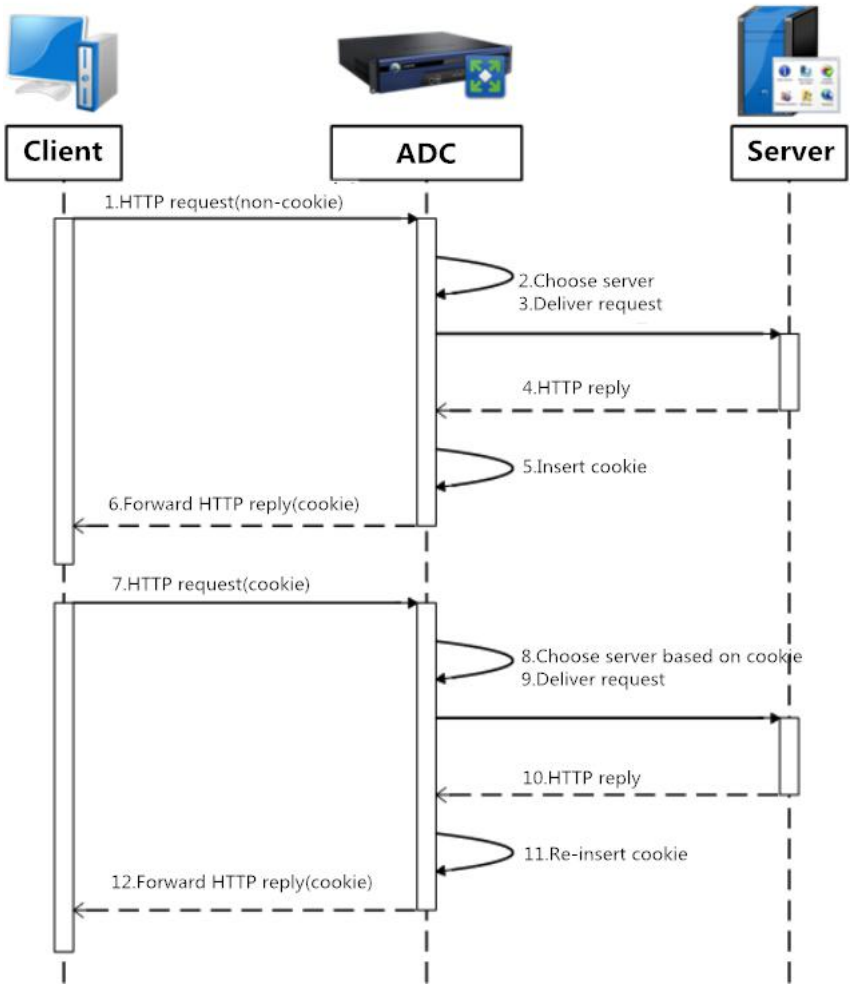
2.3.5 Session Persistence

Session persistence feature of Sangfor ADC used to distribute requests from a specific user or from the same IP to the same back-end server, so as to respond to client requests seamlessly and more efficiently, and meanwhile reduce the amount of new connections so as to reduce load of ADC.

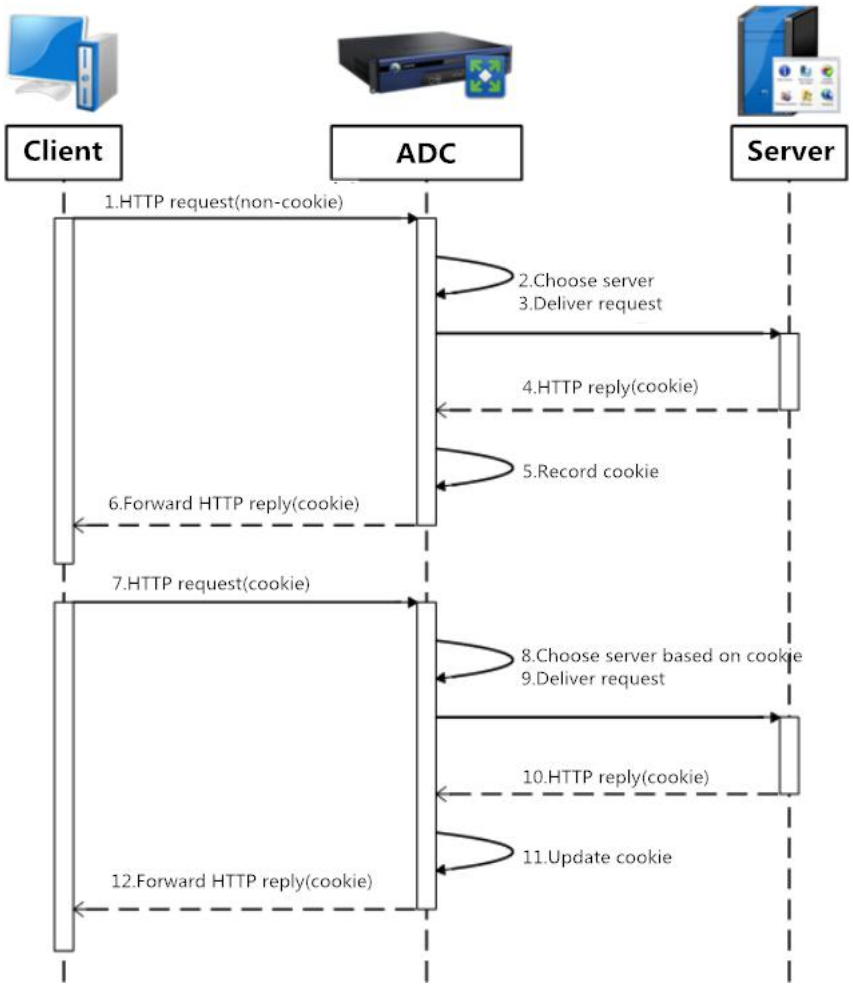
- **Session persistence based on source IP**: This feature of Sangfor ADC used to distribute requests from the same IP to the same back-end server. Another important factor is connection timeout value. Each session is configured a timeout value. If the interval between the completion of last session and start of this session is smaller than this value, this session will persist. Yet if the interval is larger than this timeout value, this session will be taken as a new one and will be scheduled to a specific server. The following figure shows detailed procedures:



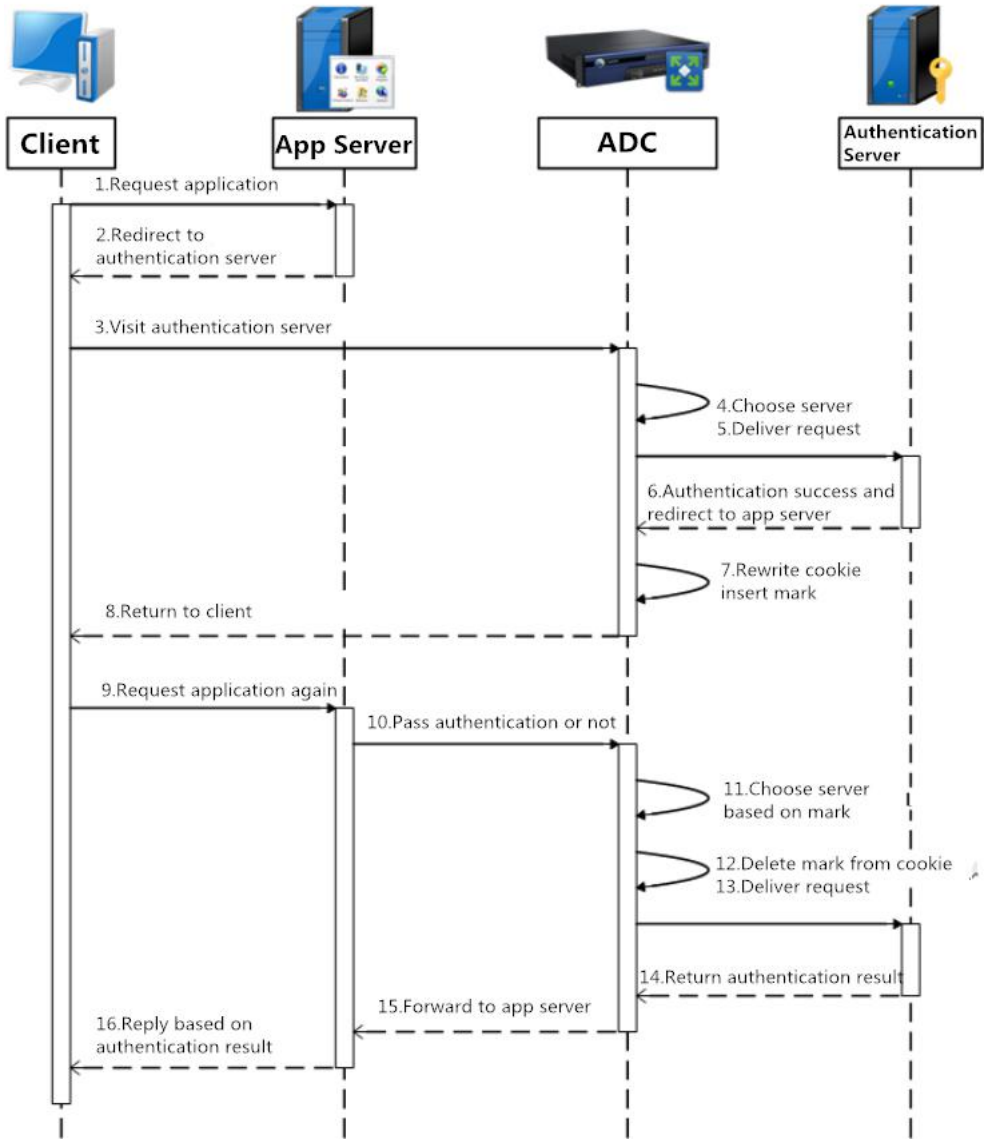
- **Session persistence based on cookie insert**: This feature is based on cookie persistence which requests according to cookie contained in client requests. While forwarding responses from server to clients, Sangfor ADC will add a special cookie to the client HTTP requests. With every request the client makes, it sends this cookie which the load balance decodes to determine which server to send the client to and requests with the same cookie will be distributed to the same server. The following figure shows detailed procedures:



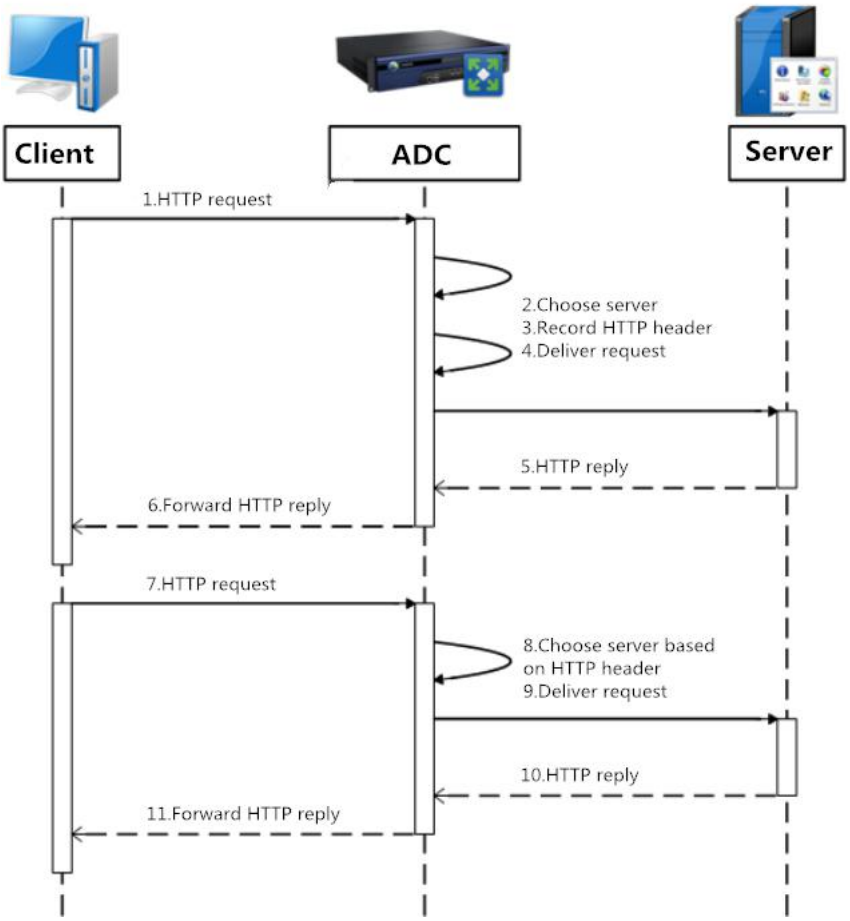
- **Session persistence based on cookie passive**: This feature supports scheduling requests based on cookie contained in client requests, which is quite similar to session persistence based on cookie insert. The difference lies in that cookie will be added to client HTTP requests by servers when servers send responses back. Sangfor ADC will remember this cookie while forwarding responses from servers to clients. With every request the client makes, it sends this cookie which the load balance decodes to determine which server to send the client to and requests with the same cookie will be distributed to the same server. The following figure shows detailed procedures:



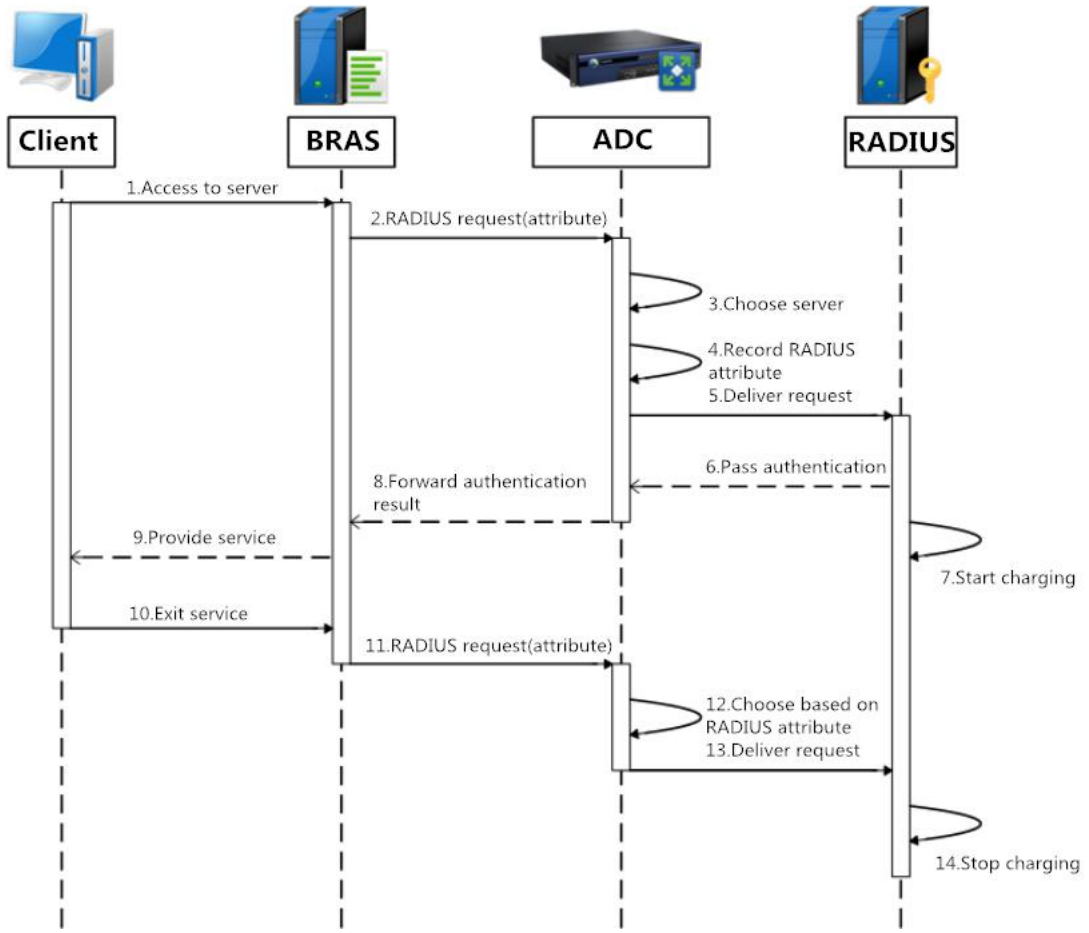
- **Session persistence based on cookie rewrite**: This feature is mostly used in the scenario where there is interaction between authentication system and application system. When first access to application server, client request will be redirected to authentication server first. Sangfor ADC will do load balance in multiple authentication servers scenario, meanwhile, it will also write identification information into cookie that is returned by the authentication server so as to differentiate authentication servers. Sangfor ADC will identify the authentication server to which client request is redirected previously based on the identification information, then it will delete identification information from cookie and forward authentication request, so as to realize session persistence. The following figure shows detailed procedures:



- **Session persistence based on HTTP header**: Some HTTP applications does not support cookie-based session persistence, but session information can be contained in HTTP header. For example, when mobile users gain Internet accesses using WAP, Call-ID is used to represent mobile phone number of the mobile user. In this case, Sangfor ADC will remember session information contained in the HTTP header. Every request which contains the same session information will be scheduled to the same server so as to realize session persistence. The following figure shows detailed procedures:



- **Session persistence based on RADIUS**: RADIUS requests of broadband users are sent to RADIUS authentication server through BRAS(broadband remote access server). When Sangfor ADC load balance for RADIUC authentication servers, it will also maintain the consistency of sessions based on RADIUS attribute. All connections associated with one user session will be directed to the same RADIUS server, so that consistency of authentication and charging of RADIUS server can be ensured. The following figure shows detailed procedures:



2.3.6 Server Health Check

Sangfor ADC supports health check based on hardware status, application type and observations, and customized health check is also supported.

- **Proactive health check based on hardware status**-Check real-time status of servers by executing ping command or based on SNMP. If the following events occur, for example, no packet is returned to ping command, resource consumption of server is too high, or system is stuck, etc., client requests will be distributed to another normal server.
- **Proactive health check based on application type**-Check real-time status of applications based on application types, be it HTTP, FTP, E-mail, DNS, RADIUS, etc. Should failures occur, client requests will be distributed to another normal server.
- **Passive health check based on observations**-Check real-time status of servers by observing such factors as sessions, inbound and outbound data packets, etc. For example, Sangfor ADC detects such errors as HTTP 403 Forbidden error, HTTP 404 error, etc, detects abnormal data transmission using TCP, for example, connection termination caused by RST flag, TCP zero window, etc. Should such circumstances occur, client requests will be distributed to another normal server.

- **Custom**-Sangfor ADC can check whether the server is normal based on the pre-defined customized character string. For example, Sangfor ADC will check whether the data packets returned by the server contain the character string which has been previously configured, if not, Sangfor ADC will take is as a failed server and will direct client requests to another server.

2.3.7 Server Exit (session persistence traffic allowed only)

When a certain server should be maintained or upgraded, Sangfor ADC ensures that client requests will not be interrupted when the server is exit from the server list. Once a server is selected to be exit, it will not be maintained or upgraded until it completes processing all the current requests, and Sangfor ADC will not direct new client requests to the server.

2.3.8 Server Warm-up

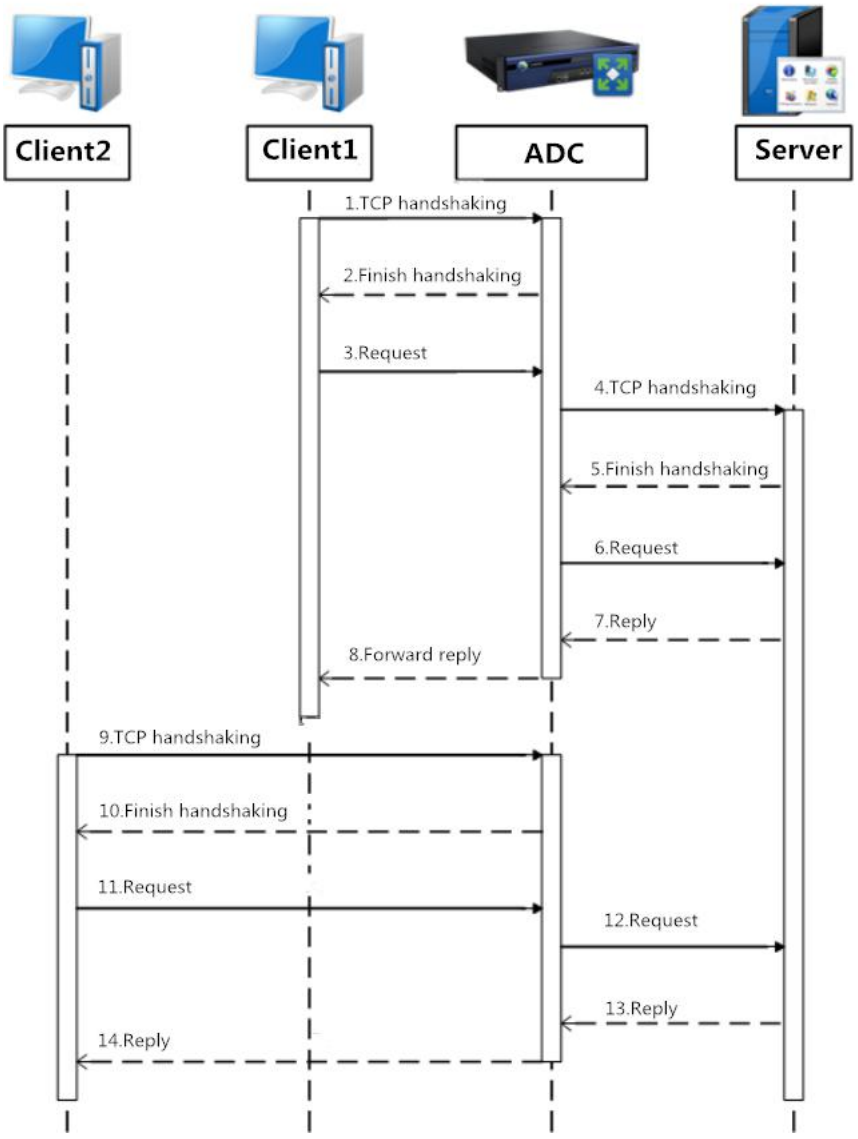
When new servers have been maintained are added to server list, Sangfor ADC can guarantee operation of servers by preventing too much traffic load on those servers. Sangfor ADC will not direct client requests to newly added servers during server recovery time, and will gradually increase requests to the server, so as to ensure that server can operate normally during the warm-up period.

2.4 Server Performance Optimization

Informatization is a long-term process for all the enterprises or organization, and hardware investment and performance are too factors which are of great concern to IT departments. Compared with traditional load balances, Sangfor ADC not only provide load balance function to improve utilization of server resources, but also provide multiple performance optimization technologies such as TCP connection reuse, memory caching, HTTP compression, SSL offloading, etc. Sangfor ADC helps to reduce investment in hardware resources, reduce server response time, improve speed and stability.

2.4.1 TCP Connection Reuse

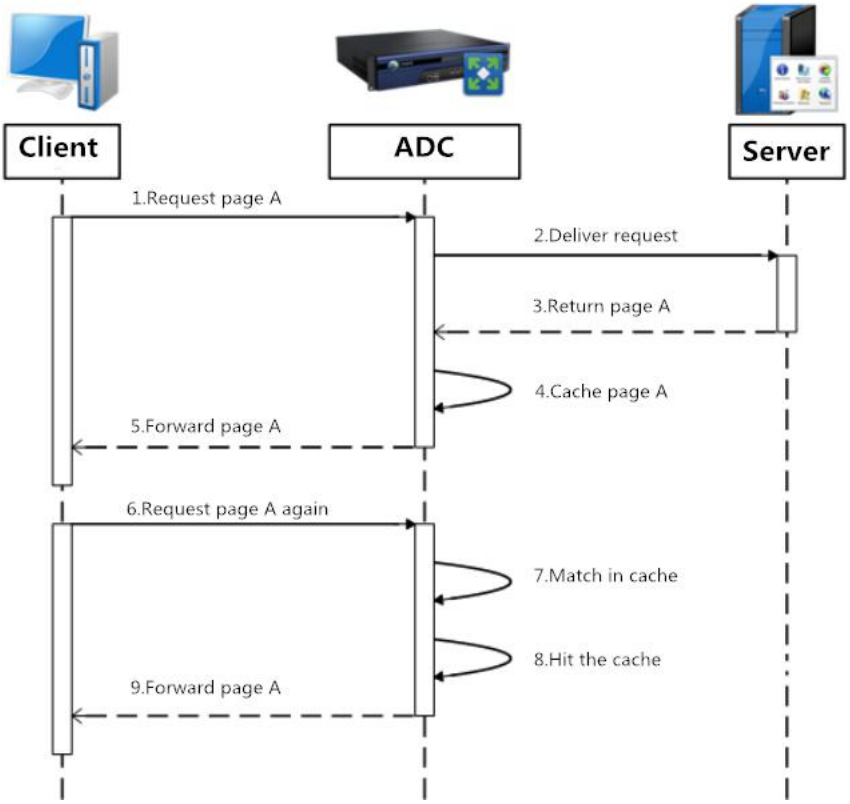
Sangfor ADC supports directing multiple client requests to the same TCP connection which is not frequently used. There is no need to establish a new TCP connection for every single client request. Therefore, server load can be reduced and processing capability of servers can be improved, without changing network topology or increasing hardware investment.



Sangfor ADC keeps the TCP connection which has been established before, greatly reducing the amount of client requests to be processed by servers(up to 90% at most), speeding up processing between client and server, improving processing capability of the application system, and reducing hardware investment.

2.4.2 RAM Caching

Based on the reverse proxy cache, Sangfor ADC supports dynamically adjusting cache space and it is much faster than other products.

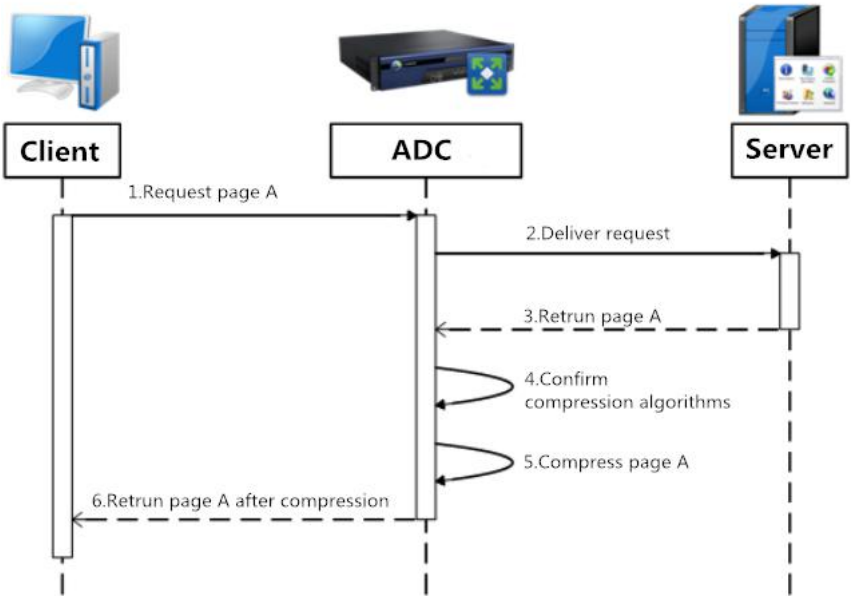


With RAM caching, Sangfor ADC helps to reduce server load, hardware investment, meanwhile, improve system processing capability and user experience.

2.4.3 HTTP Compression

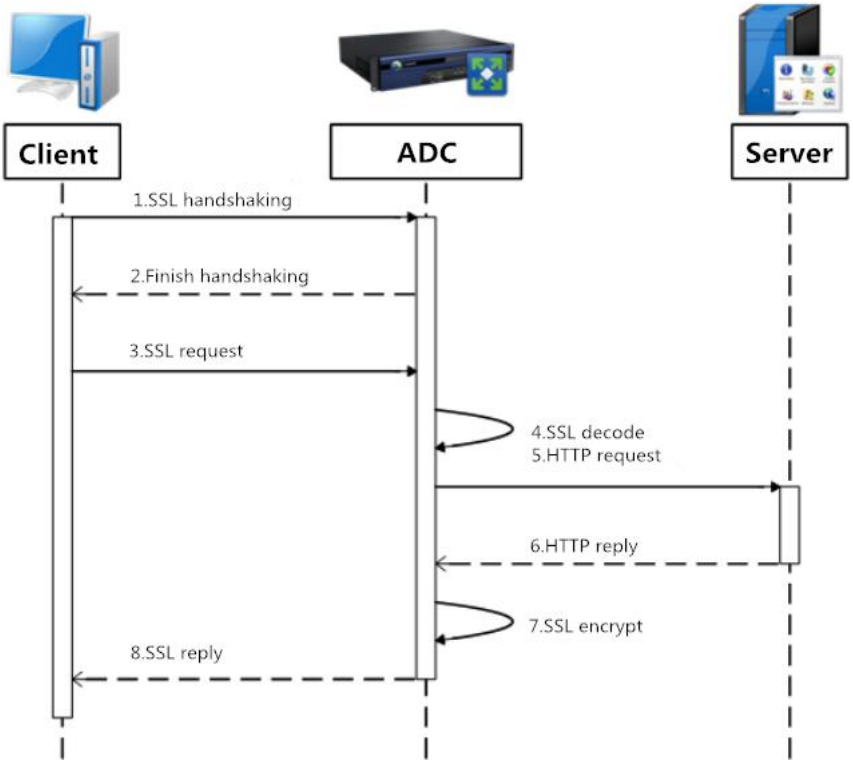
Sangfor ADC support identify whether the compression scheme gzip or Deflate is supported by the client side and supports compressing data using HTTP compression scheme gzip and Deflate dynamically.

Sangfor ADC helps to save bandwidth and improve bandwidth utilization, reduce pressure on web servers, reduce hardware investment, improve download speed, and most of all, improve user experience.



2.4.4 SSL Off Load

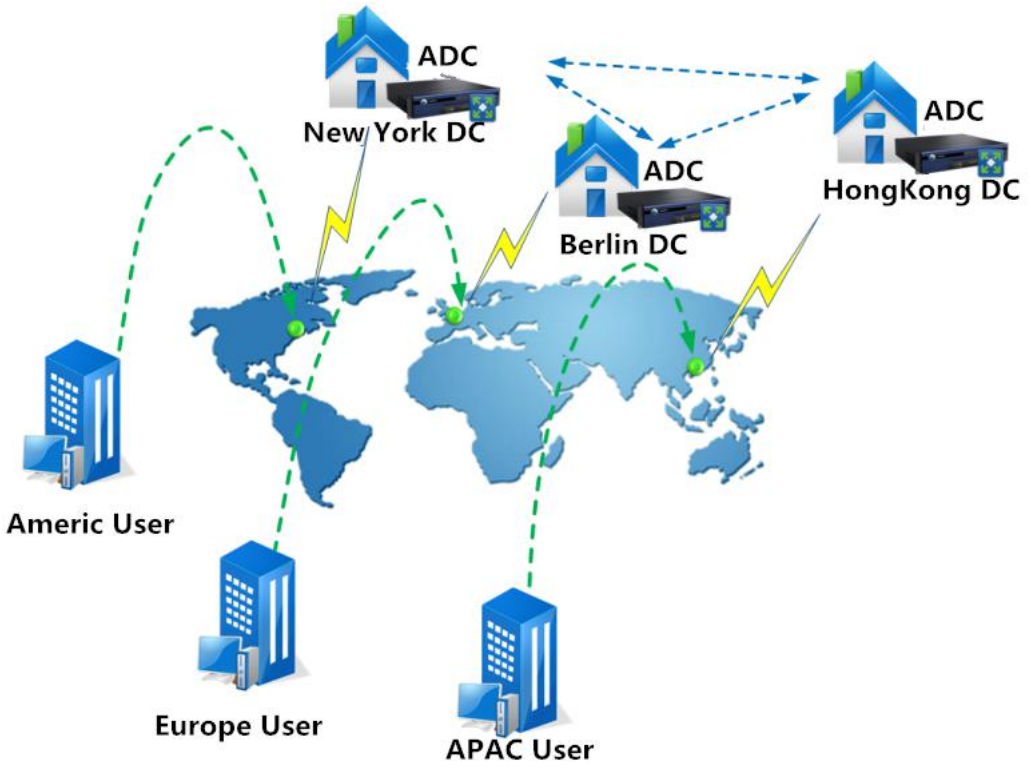
SSL off load on Sangfor ADC relieves a Web server's processing burden of encrypt and decrypt traffic sent via SSL, so as to improve response speed. Sangfor ADC supports encrypting traffic sent via SSL from server to client, configuring encryption algorithms, and managing server certificates.



Sangfor ADC helps to reduce consumption of server performance and reduce hardware investment due to the face that the amount of application system servers can be reduced, meanwhile, it also helps to greatly reduce response time and therefore drastically improve user experience.

2.5 Global Server Load Balance(GSLB)

Sangfor Global Server Load balance supports distributing traffic among different data centers located at different locations of the planet, improve service continuity, high availability and responsiveness, avoiding local downtime, and providing faster and more stable user experience.



Based on comprehensive health check mechanism, Sangfor ADC supports monitoring health and responsiveness of data centers, detecting failed data centers or servers timely, and directing successive requests to other normal data centers or servers. It not only helps to realize redundancy among different sites, but also improve access stability and resource utilization of the sites.

2.5.1 Multiple-site Scheduling based on Intelligent DNS

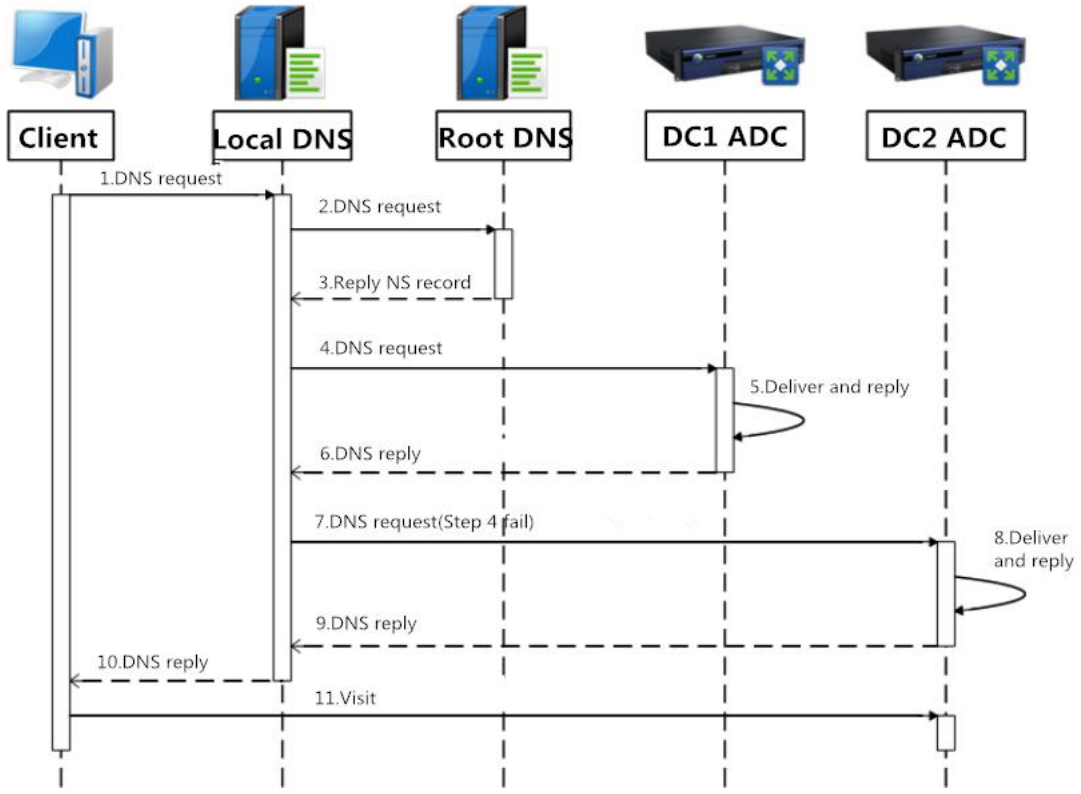
1. Realization

Based on intelligent DNS, data centers which deliver the same service will be provided with the same interface and client requests will be directed to different data centers based on the load balance policies that have been configured by administrators, so as to realize load balance among different sites.

By calculating proximity based on location of local DNS, an IP address of an optimal site will be resolved when user requests for server resources with domain name. Meanwhile, the global IP address database of Sangfor ADC helps to improve accuracy of allocation of requests based on proximity, so as to solve the problem of slow transfer speed because of location and different ISP.

2. Work Flow

Work flow of multiple-site scheduling based on intelligent DNS:



Detailed description of the work flow:

Steps	Description
1	Client of WAN user sends DNS request to local DNS server.
2	Local DNS server searches whether there are relevant records locally. If not, it will send queries to root DNS server.
3	Root DNS server responds to local DNS server with two NS records, indicating DC1 and DC2.
4	Local DNS server sends DNS requests to Sangfor ADC of DC1.
5	Sangfor ADC of DC1 checks link status first, and then select an appropriate IP address based on the pre-defined load balance algorithm.
6	Sangfor ADC of DC1 sends the IP address to local DNS server.
7	Local DNS server fail to receive DNS results sent back from Sangfor ADC of DC1 and sends DNS requests to Sangfor ADC of DC2.

8	Sangfor ADC of DC2 checks link status first, and then select an appropriate IP address based on the pre-defined load balance algorithm.
9	Sangfor ADC of DC2 sends the IP address to local DNS server.
10	Local DNS server forwards the IP address to client.
11	Client sends request with the IP address. The request will finally be directed to DC2.

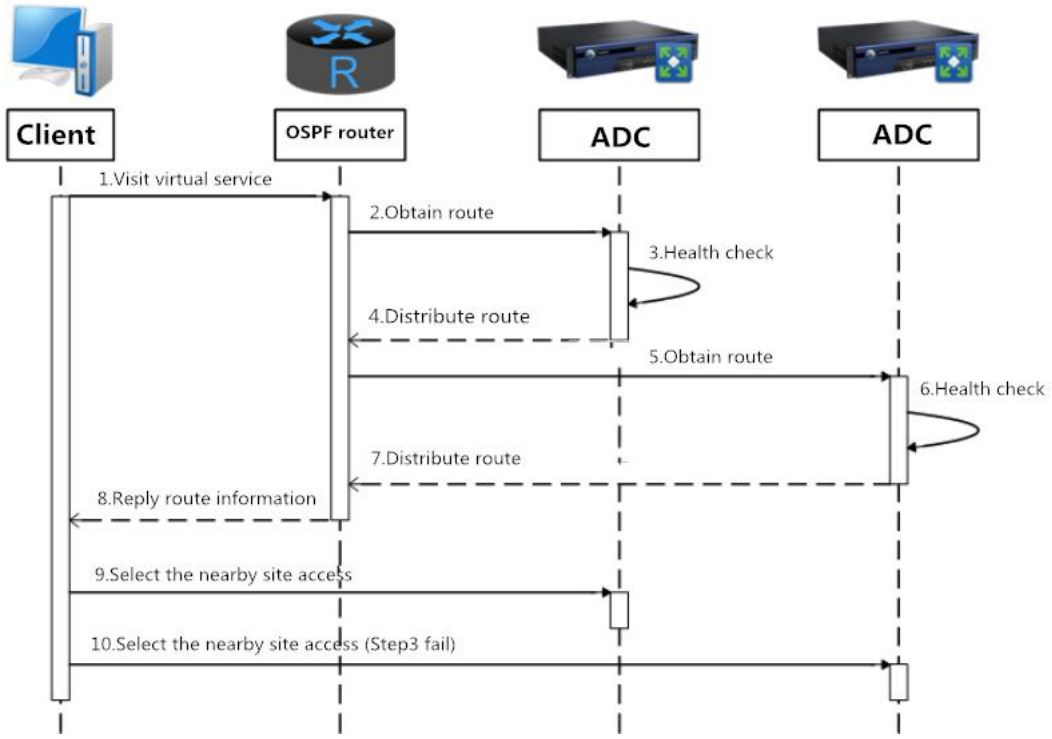
2.5.2 Multiple-site Scheduling based on IP-Anycast

1. Realization

Based on OSPF, Sangfor ADC supports site redundancy and directing client requests to the nearest available site based on routing information. When there are multiple sites, client requests will be directed to the nearest site based on the routing information. When a site fails, route to the site will be deleted by OSPF router and the site will not be selected any more. After the site recovers to normal state, route to the site will be added again by OSPF router. The speed of route notification depends on scale of network. Under normal circumstances, it takes less than 1 minute.

2. Work Flow

Work flow of multiple-site scheduling based on IP-Anycast:



Detailed description of the work flow:

Steps	Description

1	WAN client sends requests for virtual service.
2	OSPF router obtains dynamic routing information of DC1 from Sangfor ADC in DC1.
3	Sangfor ADC in DC1 checks health status of virtual service. If the virtual service is available, Sangfor ADC will distribute route to the virtual service.
4	OSPF router adds the route information distributed by Sangfor ADC in DC1.
5	OSPF router obtains dynamic routing information of DC1 from Sangfor ADC in DC2.
6	Sangfor ADC in DC2 checks health status of virtual service. If the virtual service is available, Sangfor ADC will distribute route to the virtual service.
7	OSPF router adds the route information distributed by Sangfor ADC in DC2.
8	OSPF router sends dynamic routing of the virtual service to client.
9	Client select the nearest site based on the route information. Client request will finally be directed to DC1.
10	When Sangfor ADC in DC1 detects that virtual service is not available, the route to the virtual service will be deleted by OSPF router, and when client selects the nearest site based on the route information, client request will finally be directed to DC1.

2.5.3 Proximity

In order to ensure that client request will be directed to an optimal data center, global loading balances should take such factors as distance among different sites, delay, load of the data center, etc., into consideration and select an optimal data center. Sangfor ADC supports both static proximity and dynamic proximity, which can be used together.

- **Static Proximity** - Sangfor ADC has a global IP address database and supports updating the database to the latest. If the destination IP address belongs to certain ISP or in certain area, Sangfor ADC will direct client request to the data center or link of that ISP or in that area. If the destination IP address is not in the global IP address database, Sangfor ADC will first find out area or ISP to which this IP address belong and then select an optimal data center based on static proximity. If it is still unable to determine which area or ISP the IP address that is requested by the client belong to, use dynamic proximity.
- **Dynamic Proximity** - Based on transmission delay of different data centers and real-time load of links of data centers, Sangfor ADC will select an optimal route and direct client requests to an optimal data center.

2.5.4 Health Check

Sangfor ADC supports checking real-time status of link and virtual service, so as to ensure that client request will be directed to an optimal site.

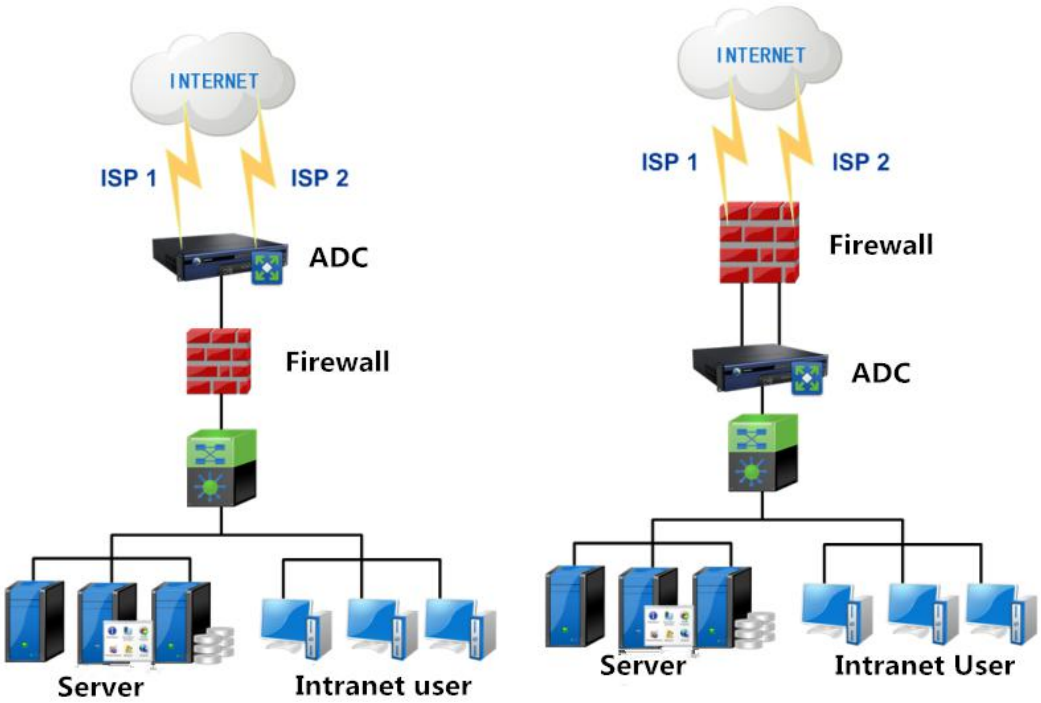
- **Link Health Check** - Sangfor ADC supports checking status of one link by checking accessibility of multiple websites. For example, check whether www.google.com and www.facebook.com are reachable using ISP1, and execute OR operation against the results. Therefore, as long as one of those sites is reachable, the link is working properly. This method avoids the limitations of checking using ICMP and also avoids mistakes that may be brought about by checking a single site.
- **Virtual Service Health Check** - Every Sangfor ADC will check health status of virtual services of all the data centers. Therefore, failed data centers can be found timely, and status of virtual services on all the seven protocol layers can be monitored. If Sangfor ADC detects that certain data center or server fails, then the client requests will be automatically directed to another normal data center or server.

2.6 Deployment and Management

Sangfor ADC can be deployed as Route mode and Bypass mode. It supports WAN protocols such as VLAN, STP, etc, and works well with dynamic routing protocols such as OSPF, RIPv1, RIPv2, etc, user can use it to cope with all sorts of complex network environment. It can improve such aspects as business sustainability, endpoint user experience, data center accessibility, etc.
Once SNMP is enabled on Sangfor ADC, users can search for such information as CPU usage, memory usage, new connections, concurrent connections, throughput, etc., on SNMP monitoring software, enabling users to monitor real-time hardware load of Sangfor ADC. Moreover, Sangfor ADC supports ACL, enabling users to precisely control accesses based on 5-Tuple.

2.6.1 Route Mode

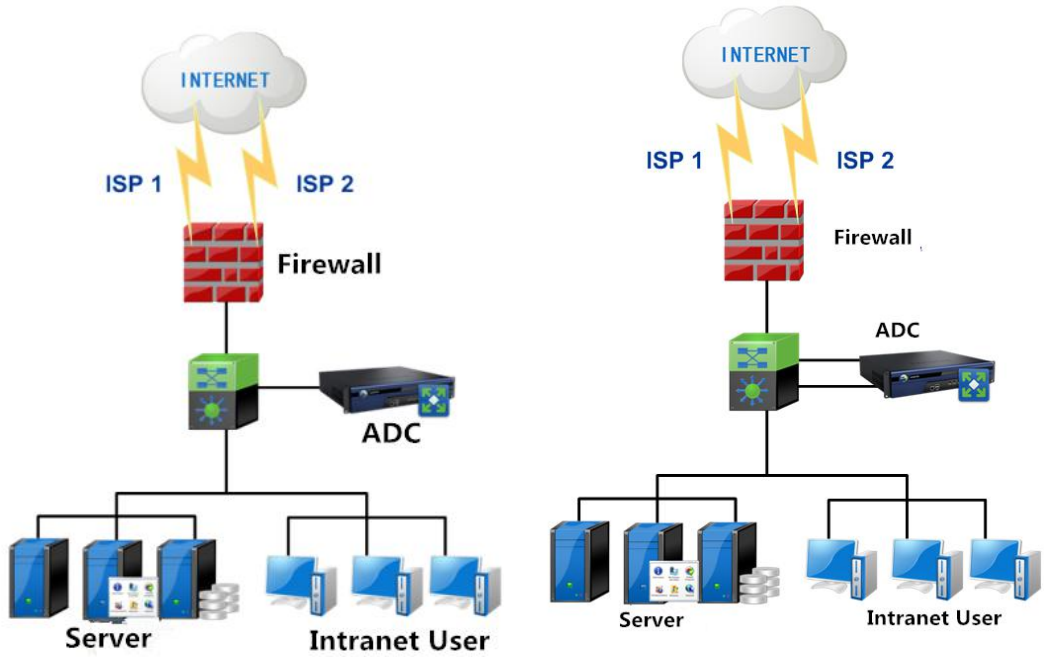
If Sangfor ADC is deployed as Route mode, all the traffic will go through Sangfor ADC and when client requests reach Sangfor ADC, Sangfor ADC will direct those requests to an optimal link and will select a server with best performance in the server group based on the pre-defined load balance policies, ensuring user experience and improving user satisfaction. This deployment mode is applicable to such scenarios as link load balance, server load balance and global load balance.



- **Route Mode Ⅰ** -This is the most common deployment mode of Sangfor ADC. It supports server load balance and multiple-link load balance. NAT should be implemented on ADC and firewall should be in transparent mode.
- **Route Mode Ⅱ** -NAT(WAN1-LAN1 / WAN2-LAN2) should be implemented on firewall and ADC functions as a router. This mode supports server load balance and multiple-link load balance. Firewall should be connected with Sangfor ADC with two network cables so as to support DNS mapping.

2.6.2 Bypass Mode

This mode does not change network structure and service will not be interrupted when deploying Sangfor ADC. When client requests reach Sangfor ADC, Sangfor ADC will select an optimal server based on the pre-defined load balance policy, so as to improve server utilization and avoid uneven distribution of load on servers. This mode is applicable to server load balance.

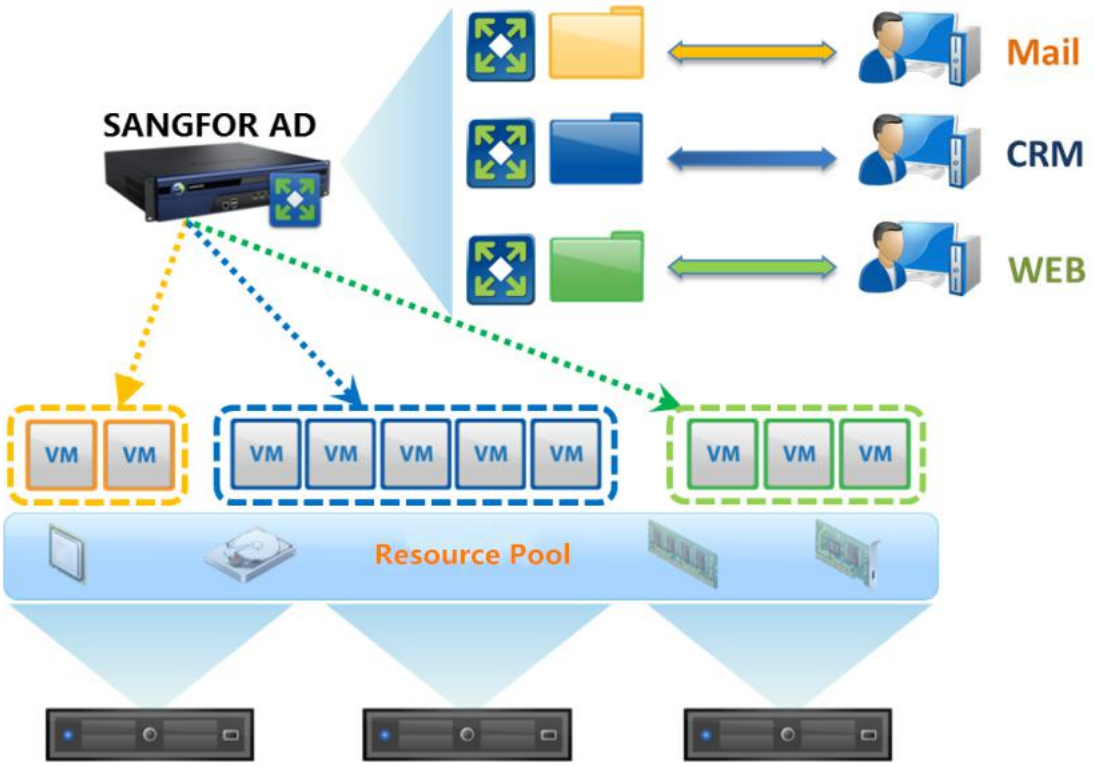


Bypass Mode Ⅰ **Bypass Mode Ⅱ**

- **Bypass Mode Ⅰ**-NAT should be implemented on firewall and traffic should be bypassed to Sangfor ADC. As long as DNS requests can reach Sangfor ADC, link load balance on inbound traffic can be realized, server load balance can be realized as well.
- **Bypass Mode Ⅱ**-Sangfor ADC should be connected with two network cables. Firewall should support policy-based routing, which should be applied to WAN interface of Sangfor ADC, so as to realize link load balance on inbound and outbound traffic, and server load balance as well.

2.6.3 Virtual Partition

One Sangfor ADC can be configured into multiple vADs, which can be allocated to multiple tenants or departments, enabling management segregation and data segregation as well as providing all the tenants or departments with full use of computing resources of Sangfor ADC.



- **Multi-tenancy -** In order to meet the requirement of multiple tenants, multiple vADs can be configured on one Sangfor ADC, which are segregated from each other. Then based on server virtualization structure, all sorts of business systems which are highly flexible and which can be easily managed can be constructed.
  - ✓ Support creating, deleting, enabling or disabling vAD based on business requirements.
  - ✓ Support allocating CPU, memory and NIC for each vAD. Support configuring new sessions ans concurrent sessions.
  - ✓ Support configuring admin password and IP address for each vAD.
  - ✓ Support VLAN division, log management and user management.

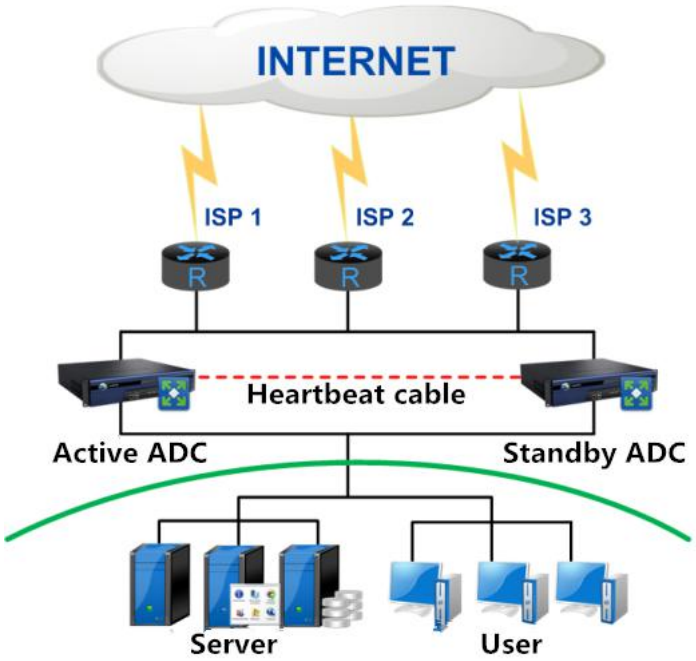
2.7 High Availability(HA)

No matter it is global load balance, or multiple link load balance and server load balance, Sangfor ADC functions as a controller whose stability and security are of paramount importance to the availability of business delivery network. In order to avoid single point of failure, HA is used to ensure business continuity and help to avoid service interruption to a large extent.

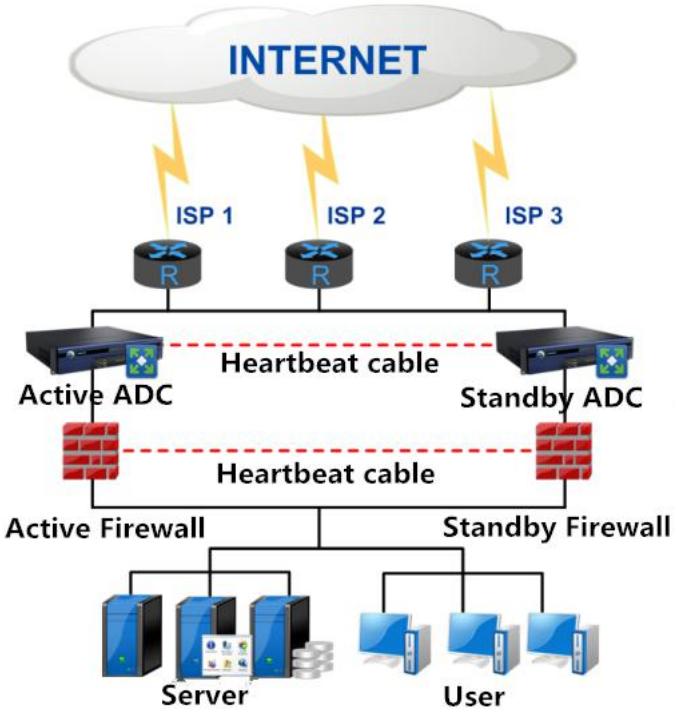
2.7.1 Active-standby

There are two Sangfor ADCs, one ADC which is performing load balance task is taken as active node and the other is taken as a standby node. When the active node is processing, it will also synchronize sessions to the standby node, so as to ensure service continuity when business is scheduled to the standby node. This mode is applicable to the scenario when the deployment requirement is to avoid single point of failure.

- **Passive Switch Over** - When the active node fails, the standby node will immediately take over the load balance task.



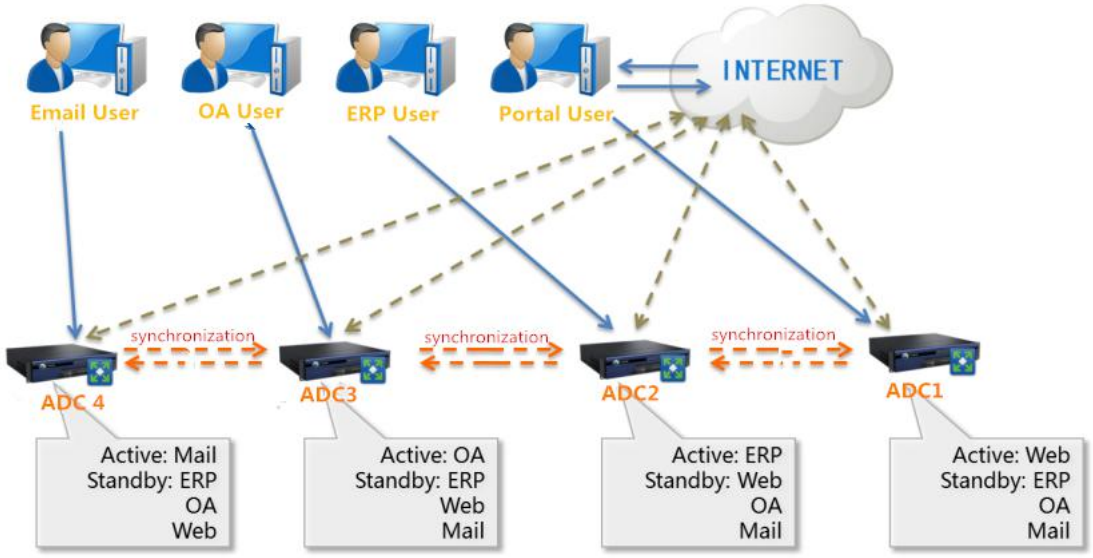
- **Active Switch Over** - When network failure is detected on the active node, for example, link health check finds that WAN interface is interrupted, ARP detection finds that active-standby switch over takes place on firewall, the standby node will be automatically changed to be an active node so as to ensure business continuity.



2.7.2 Cluster

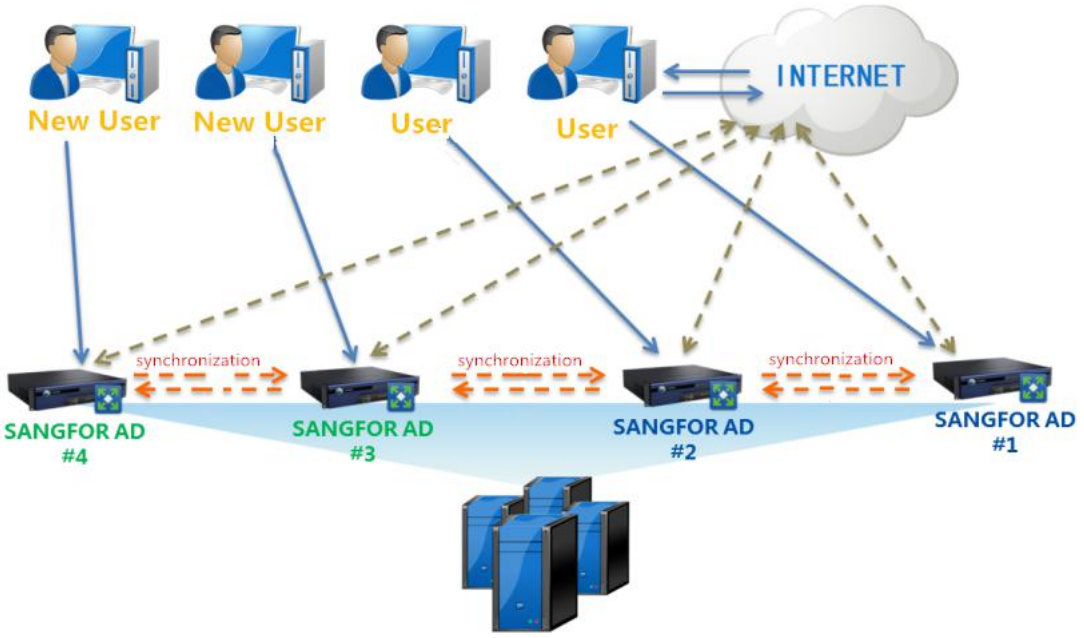
Sangfor ADC cluster helps to enhance processing performance of the overall system, make full use of every single ADC in the cluster, deliver different services, support backups and moreover reduce service interruption to the greatest extent in case of failure. No service will be interrupted as long as there is one ADC available in the cluster. A balanced and dynamic switch over in case of failure is supported.

- **HA Cluster**
  - Sangfor ADC HA cluster supports multiple modes, such as active-standby, two actives, multiple actives, M+N, etc.
  - Compared with HA, HA cluster provides better redundancy(Multiple standbys vs one standby) and lower cost(1+N vs 2×N).
  - Sangfor ADC supports mutual backups and synchronization of session mirror, ensuring service continuity.



➢ **High-performance Cluster**

- Support multiple Sangfor ADCs, which are virtualized into one logic device.
- Support performance expansion for businesses whose bandwidth is more than 100Gbps.
- Support combination with HA cluster so as to better support business development.



3. Special Features of Sangfor ADC

3.1 One-Way Acceleration

One-way acceleration used to detect and real-time monitor transmission delay, packets loss and re-transmission automatically, so as to change data transmission mechanism, relieve transmission congestion, avoid re-transmission of messages, and most of all reduce response time and improve efficiency of transmission over TCP. Since one-way acceleration helps to optimize all the TCP data flow, acceleration can be realized in file transmission, email sending and website browsing, etc. As long as the application is based on TCP, Sangfor ADC can help to speed up transmission and enhance user experience.

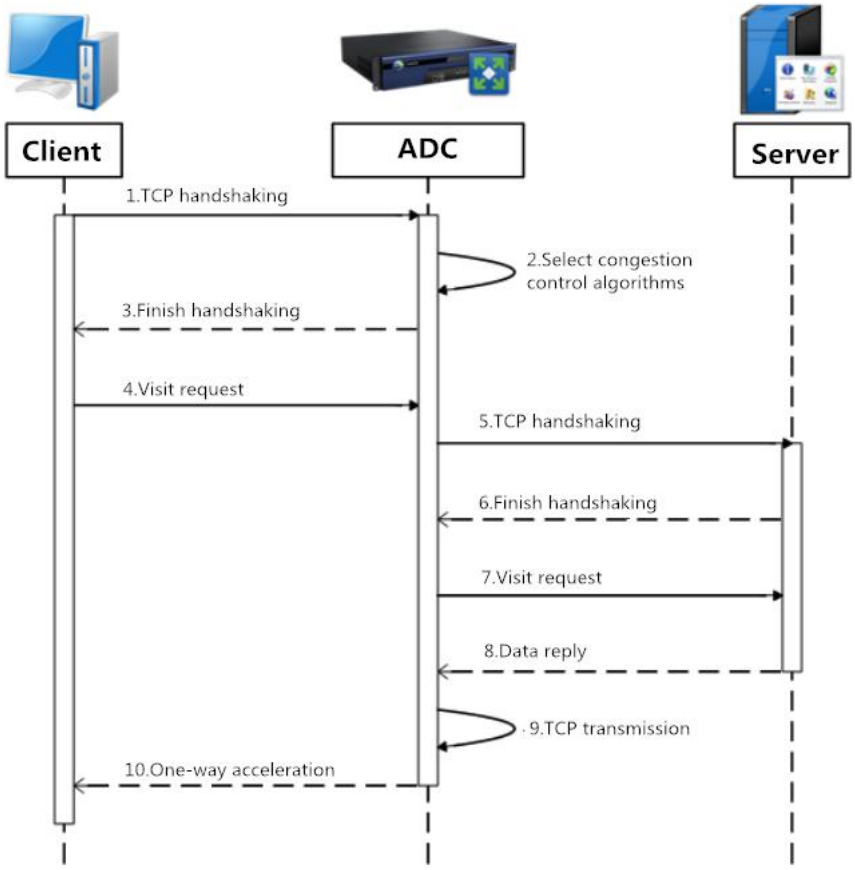
3.1.1 Background

During the process of constructing business system, attention should not only be paid to ensuring stability and continuity, but also be paid to improving access speed and improving user experience. Despite the fact that many organizations have links of multiple ISP, transmission speed is still slow when WAN users request for LAN resources, especially for stock users, 3G/4G users or users request for overseas resources. The reason is that there are problems such as transmission delay and packets loss during the transmission process, therefore, undoubtedly, transmission speed is slow.
Traditional solution focuses on improving bandwidth, deploying more links, or deploying link load balances to improve access speed and stability. But actually this solution is not effective at all, because it not only increases bandwidth cost, but also fail to improve link quality, and most of all, it does not help to reduce response time. Therefore, under most circumstances, access speed is not improved efficiently.
As for traditional WAN optimization solutions, such technologies as flow cache, data compression, protocol optimization, etc., are employed. WAN optimization helps to improve link quality to some extent, yet it requires to deploy WOC both at the client side and server side. This solution is applicable to the scenario when there are sites and CMC. But when users are scattered, this solution can not be effectively carried out. Therefore, it requires an effective and feasible solution.

3.1.2 Mechanism

Traditional TCP is based on LAN, therefore, if TCP is used in an unstable environment such as WAN, it will bring about transmission delay and packet loss and therefore bring down data transmission speed. One-way acceleration helps to optimize congestion control algorithm, address the issue of the defects of TCP, and realize transmission acceleration. The core part of one-way acceleration is optimization of congestion control algorithm, for example, slow start, congestion avoidance, fast retransmit, fast recovery, etc.

- **Congestion Avoidance** - It estimates available bandwidth precisely and knows congestion window based on estimation, so as to make most use of bandwidth.
- **Fast Re-transmit** - It allows receivers to use TCP SACK option to find out at most four non-contiguous blocks. RFC 2883 is used to find out reused fields in SACK TCP option of re-transmitted data packets. Therefore, sender can know when the unnecessary field is re-transmitted and prevent unnecessary re-transmission later. The less re-transmission, the more reasonable throughput is.
- **Fast Recovery** - It detects packet loss quickly and re-transmits the lost packets immediately, improving bandwidth usage efficiently when the situation is that transmission delay is long and network is in poor condition. By changing the method employed by sender to increase transmission speed during fast recovery, it helps to increase throughput.
- **Slow Start**-Slow start or congestion avoidance refers to the algorithms which help to avoid congestion in transmission using TCP. During the initial data sending period or when restoring lost segments, the algorithms help to increase window for sending data, i.e., clients can send more segments. Slow start algorithm helps to increase window with a complete TCP segment for each segment that has been received or each segment that has been confirmed. Congestion avoidance algorithm helps to increase window with a complete TCP segment for each segment that has been confirmed. The two algorithms help to increase transmission speed at sending window so as to make full use of bandwidth.

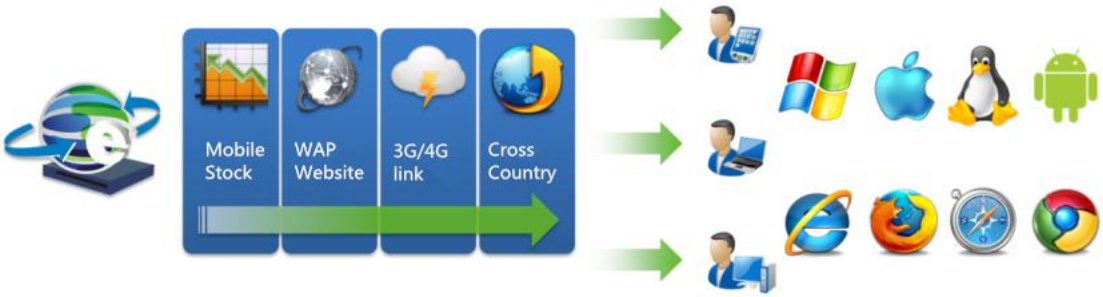


On one hand, bandwidth keeps changing, therefore, if data is transmitted too fast, congestion may be brought about, if data is transmitted too slow, bandwidth utilization may be too low. One-way acceleration will adjust the frequency of sending packets based on real-time network status, so as to

make best use of bandwidth. On the other hand, as for standard TCP algorithm, too much data will be re-transmitted in case of packet loss, resulting in low transmission speed. One-way acceleration will detect packet loss, send redundant data and reduce re-transmission rate.

3.1.3 Scenario

Unlike traditional WAN optimization solution, there is no need to deploy Sangfor ADC or install software client or browser plug-in at the client side. Besides, Sangfor ADC is compatible with all kinds of endpoint devices, operating systems, browsers, etc. It functions to reduce response time, improve user experience and enhance business competitiveness for organizations, without the need of upgrading bandwidth.



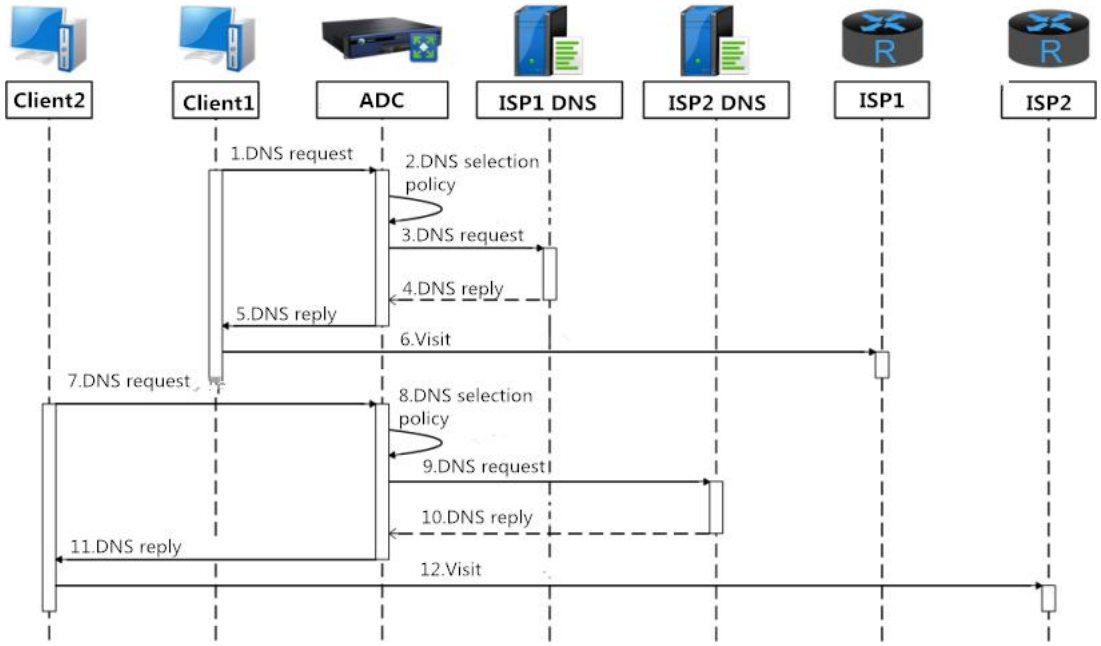
For such scenarios as stock users, 3G/4G users or users request for overseas resources, etc., one-way acceleration of Sangfor ADC helps to increase access speed by at least 30%, improve user experience, and moreover enhance productivity internally and competitiveness externally.

3.2 Intelligent Optimization

As business systems are relying more and more on network, there is also a growing demand for IT maintenance personnel. Since client requests are becoming increasingly complex, Sangfor ADC comes up with such functions as intelligent routing, transparent DNS proxy and link control, which help to increase link utilization by selecting an optimal link based on such factors as real-time link load, period, users, destination, etc.

3.2.1 Transparent DNS Proxy

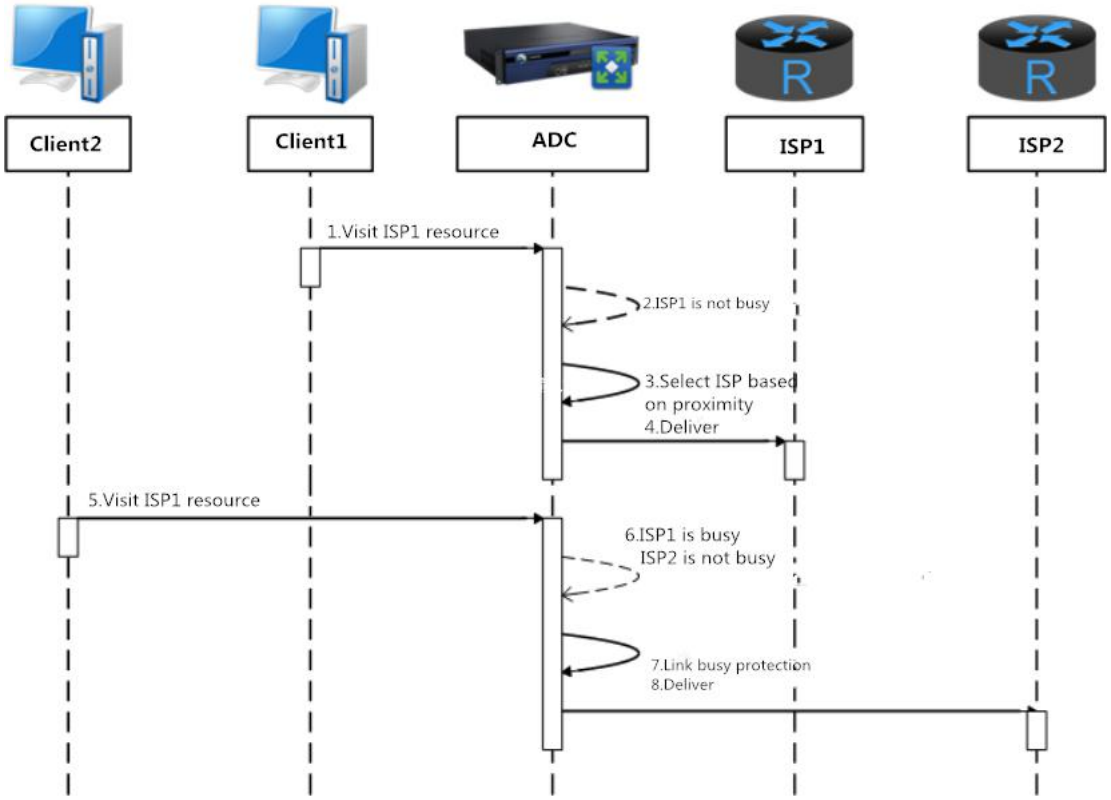
When LAN users request for WAN resources, they are required to specify one ISP's DNS server and use that ISP's DNS service. Therefore, a great amount of client requests may be directed to the same link, resulting in slower access speed because the link keeps being occupied yet the other links are not used at all. Imbalanced utilization of link may bring about waste of Internet resources, but also slower access speed.



Transparent DNS proxy of Sangfor ADC functions to forward DNS requests no matter which DNS server is specified by the clients. It will select an optimal DNS server and direct client requests to different links based on the load balance algorithm and the pre-defined link load balance policy, enabling optimal and balanced utilization of multiple links.

3.2.2 Link Busy Protection

As there is a growing demand for link stability, more and more links are used to realize redundancy, which helps to improve link stability to some extent. As for improving link utilization and increasing access speed, traditional bandwidth management devices support managing bandwidth based on applications, ordinary load balances support selecting an optimal link based on proximity. However, neither of the two types of devices support selecting an optimal link based on real-time link load. Therefore, Sangfor ADC stands out because it supports link control which helps to efficiently improve link utilization and bandwidth utilization.



Link busy protection used to optimize link scheduling based on the threshold configured for each link, together with the load balance algorithms of Sangfor ADC. When threshold of a certain link is reached, client requests will be automatically directed to another link based on the pre-defined load balance policies, so that access speed can be ensured and link resources can be preserved.

3.2.3 Elastic Load balance

As the business system is becoming more and more complex, proactive server health check is faced with some limitations. In particular, when multiple application systems are involved, proactive health check is not effective at all in finding out the failure. Moreover, proactive health check is not effective for applications based on unknown protocols and private protocols. Compared with traditional solutions, Sangfor ADC provides a flexible workload control and scheduling mechanism which can check server validity by monitoring TCP abnormal connections, and will turn on overload protection for servers that are low in performance, so as to ensure business continuity and availability.

- **Intelligent Monitoring** - Sangfor ADC checks server validity by continuously monitoring TCP traffic. Client requests will not be directed to the server if the server is taken as invalid when it meets the conditions for invalid servers.

- **Overload Protection** - When servers are low in performance, current sessions will be persisted but new sessions will not be assigned to the server until server performance returns to normal.

3.2.4 Policy-based routing

Policy-based routing enables administrators to easily configure load balance policies based on the configuration wizard even if administrators are not familiar with load balance algorithms or policies. For example, when there are multiple links, specific links are for specific businesses so as to ensure utilization of each link, and intelligent routing helps to select links based on the destination domain. It also supports load balance based on period, i.e., different load balance policies will be employed during different periods, so as to increase utilization of network and server resources to the greatest extent.

3.2.5 Intelligent Alarming

Generally speaking, users of load balances have extremely high requirement for stability of business system, and always pay close attention to real-time network status, so as to avoid potential risks and avoid huge losses that may be brought about by service interruption. Once network or business fails, Sangfor ADC will immediately send SMS messages or emails to administrators. Alarms will be given when failures occur to link, server, virtual service, failover , or when cyber attacks take place, etc.

Address: iPark, A1 buiding, Xueyuan Bvld 1001,
Nanshan District, Shenzhen, Guangdong, PRC

Tel: +60 12711 7129 (7511)
E-mail: tech.support@sangfor.com