Application Delivery Product Technology White Paper

2015 V2

Venusense

2015 V2

Abbreviations	
ADN	Application Delivery Network
ADC	Application Delivery Controller
SLB	Server Load Balancing
LLB	Link Load Balancing
GSLB	Global Server Load Balancing
DSR	Direct Server Return
RTT	Round Trip Time
SNAT	Source NAT
SNMP	Simple Network Management Protocol
SSL	Secure Socket Layer
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
VRRP	Virtual Router Redundancy Protocol
HA	High Availability
DNS	Domain Name Server
LDNS	Local DNS

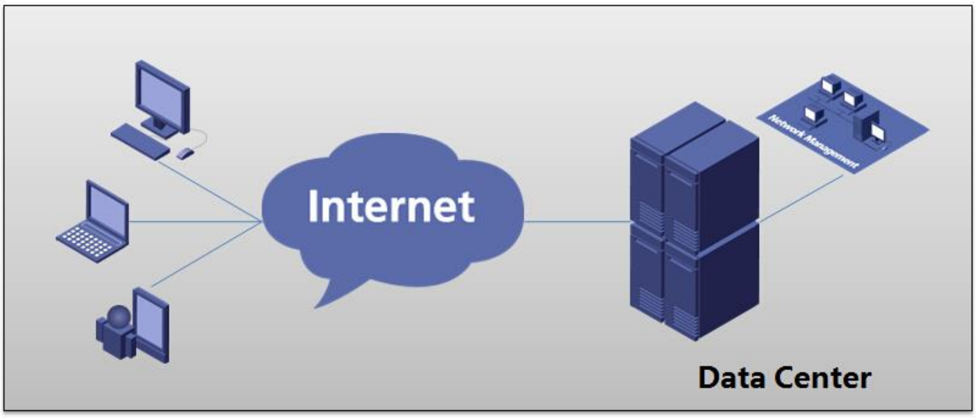
Table of Contents

Chapter 1 Concepts and Core Values for Application Delivery Product

1.1 Overview of ADN (Application Delivery Network)

Internet is essentially a kind of end-to-end technology. In fact, any complex application will eventually be attributed to data exchange between Client and Server, but its processes that go through the Internet are complex. For ISP of the application, the service will not run normally or result in low efficiency if any of these processes are not handled properly. For example:

- Problems on application performance bottlenecks, stability and scalability;
- Problems on network delay arising In case of route bottleneck While accessing different ISP;
- Problems on reasonable distribution and access efficiency when broadband user and narrowband users (mobile terminals) coexists:
- Problems on network attack and related security issues of the application system;
- Problems on the overall operating efficiency and user experience improvement of application system.



Application Delivery Networks (ADN) is a set of technical system that used to guarantee stable and efficient data exchange between Client and Server, which provides user service through Web-Based browser and Internet. According to the explanation from Gartner, the world's leading IT research and advisory company, ADN products mainly consists of two fields—WAN Optimization Controller (WOC) and Application Delivery Controller (ADC). WAN Optimization Controller products are primarily used to realize data compression between multiple data centers or branch offices and headquarters, thus improving transmission efficiency and saving bandwidth costs.

ADC products derive from Load balancing product, With load balancing products, it can associate multiple servers that provide same service via providing unique access IP address (Virtual Service) externally and providing address pool internally, Therefore, the incoming traffic can be distributed to these servers according to pre-defined policy, while the status of these servers will be monitored. If one of the servers crashed, the traffic can be reallocated to the other normal server. The ISP can always increase or decrease the number of servers in this pool to meet the needs of the service change,in this way the availability and elastic expansion in the WEB server side is established..

In addition to load balancing, ADC takes great concern about all aspects of the entire application delivery, including Client side, Server side and overall efficiency of data exchange through the network nodes. ADC can realize traffic balancing in a more specific manner based on the access content by users, as well as exchange content according to the user's own information such as browser type, Cookie and other L7's message. ADC can identify applications, as well as accelerate and optimize the process. Meanwhile, ADC can also perform server pressure Offload , including SSL protocol, content compression, Cache, and TCP connection, so that server resources can be used in their own service system with higher priority, thus enhancing the efficiency of the entire system.

1.2 Core values of application delivery product

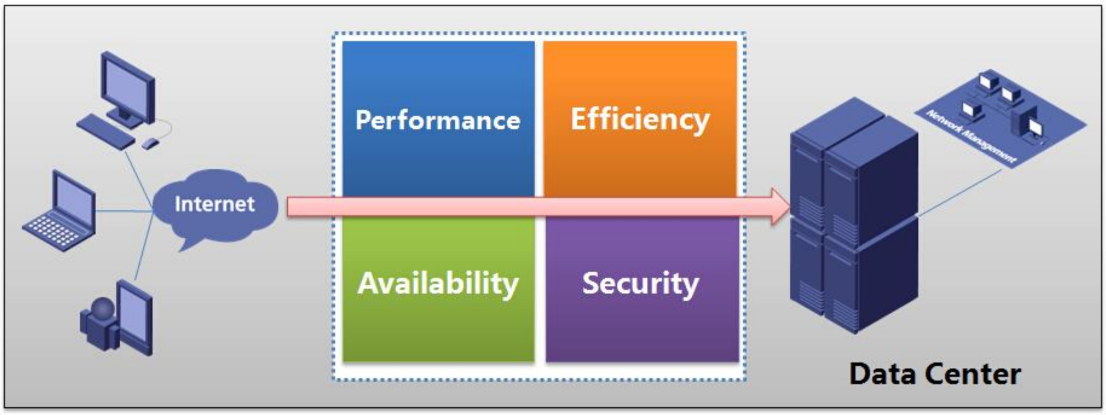
Application Delivery product focuses on all aspects of entire application delivery, while the core value aims to protect the application's following elements:

High performance --- Meet the needs of business development, and have sufficient elasticity to expand.

High efficiency --- Offload server pressure and enhances the efficiency of the entire service system.

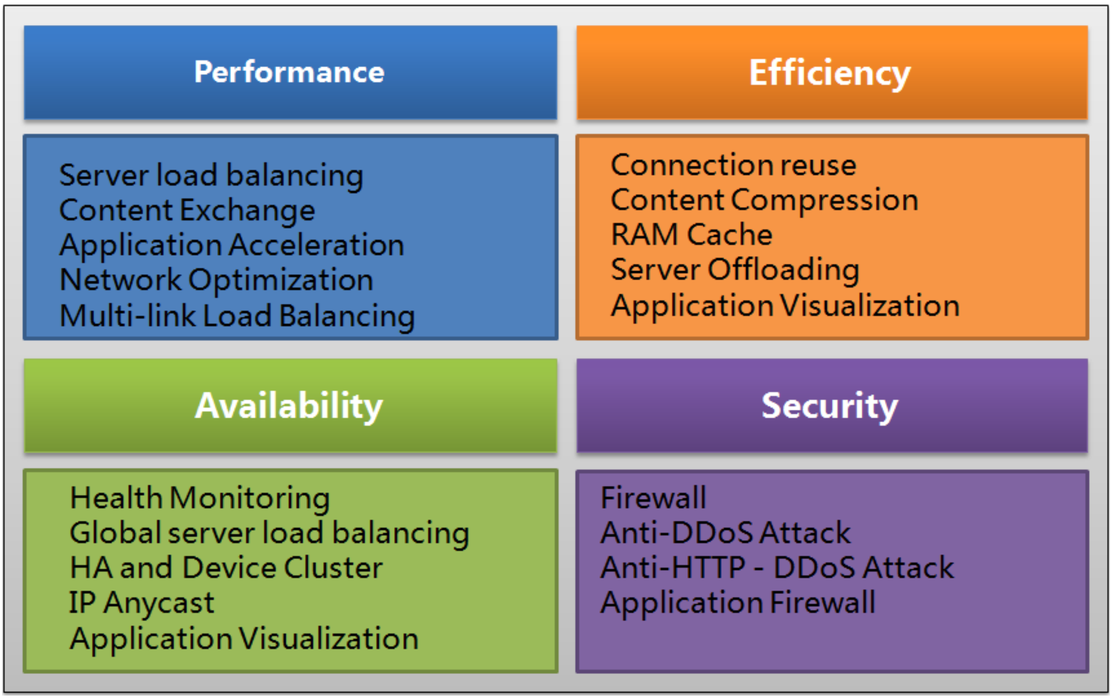
High availability---service backup and redundancy. Guarantee uninterrupted and stable operation of the service, and improve the user experience.

Security---Guarantee the service security to prevent intrusion and data leakage.



Chapter 2 Venusense Application Delivery Solution Overview

The development of Venusense Application Delivery Overview is based on strong technical accumulation,constantly exploration, the professional and effective ADC solutions are not only guarantee the continuous and reliable operation of the user application, but also save user’s investment and it brings better user experience.



Performance

Server load balancing
Content Exchange
Application Acceleration
Network Optimization
Multi-link Load Balancing

Efficiency

Connection reuse
Content Compression
RAM Cache
Server Offloading
Application Visualization

Availability

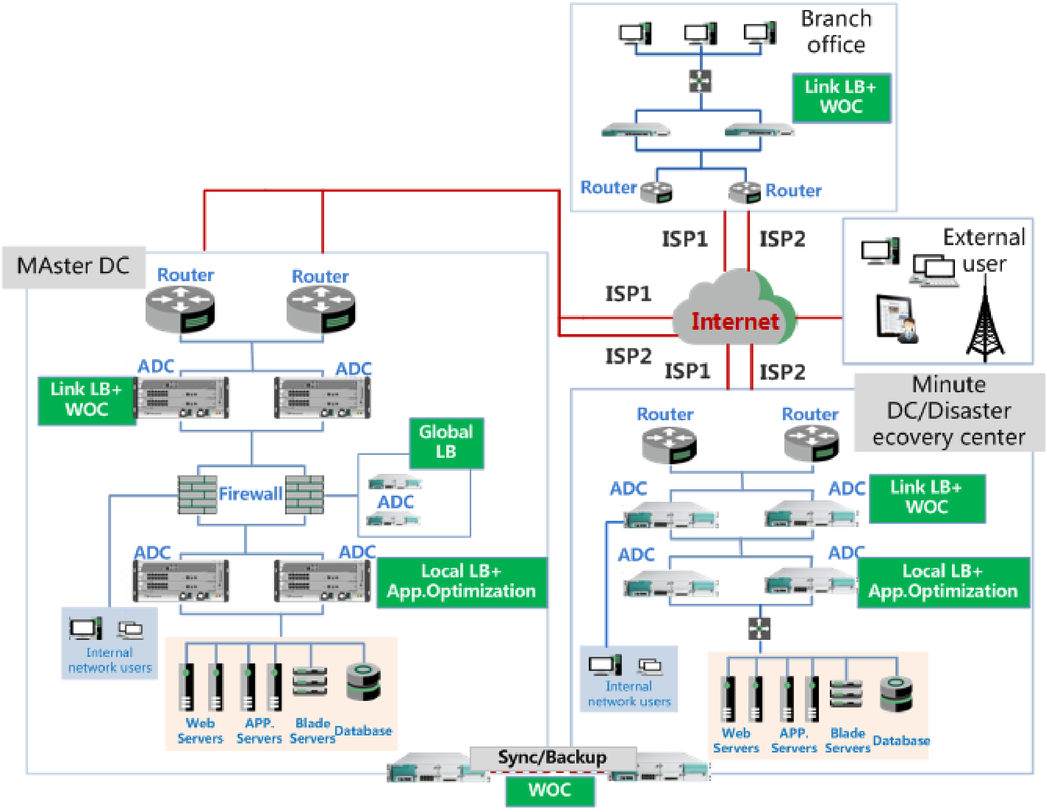
Health Monitoring
Global server load balancing
HA and Device Cluster
IP Anycast
Application Visualization

Security

Firewall
Anti-DDoS Attack
Anti-HTTP - DDoS Attack
Application Firewall

2.1 Solution Consisting

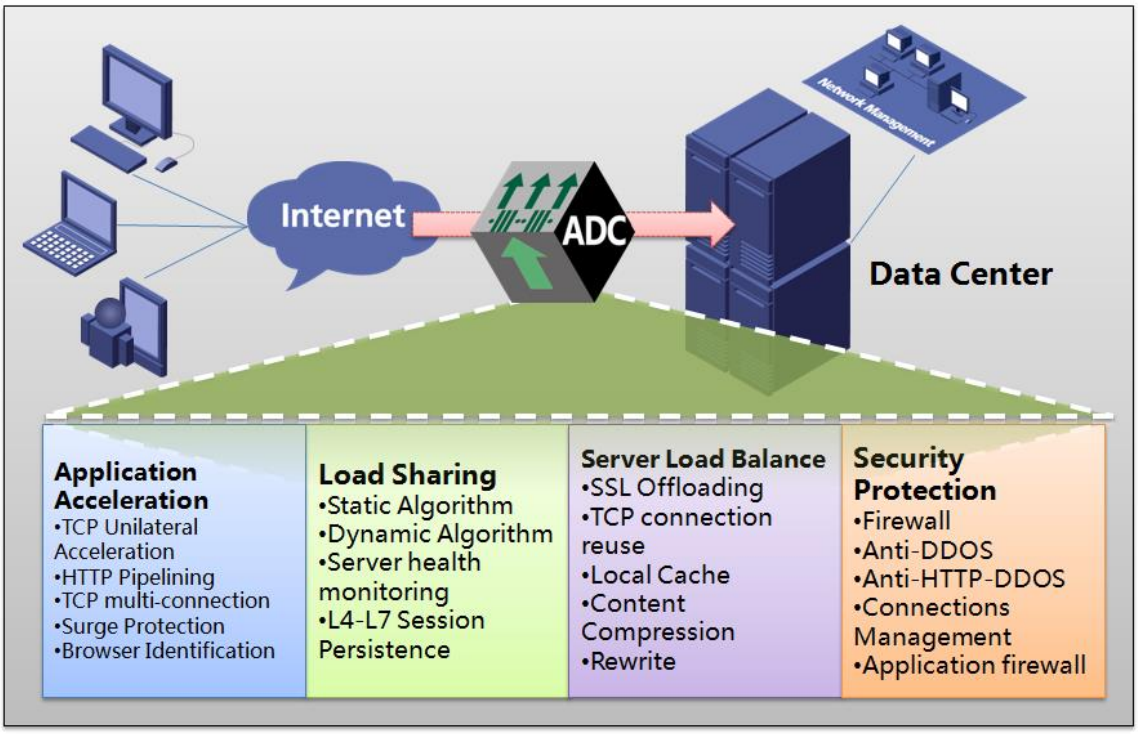
In order to meet the effective, fast and secure delivery requirement of application, Venustech launched integrated solution for Venusense ADC,including: server load balancing, application optimization and acceleration, link load balancing, global server load balancing, WAN optimization and security products ,etc.



2.1.1 Server Load balancing

Server load balancing is mainly used to manage the the access and feedback traffic of the server. Venusense ADC will intelligently distribute the access traffic to the optimal server through various static and dynamic load balancing algorithm. Meanwhile, Venusense ADC will take advantage of its high-performance multi-core hardware platforms and TBOS software systems to realise traffic compression , cache perform hardware encryption and decryption in real time , meanwhile take over the task which requires a large amount of server’s computing resource, thus making the server focus on its own task processing to achieve the improvement and efficiency enhancement of the entire system.

Venusense server load balancing solution consists of load balancing, application acceleration, server offloading etc.

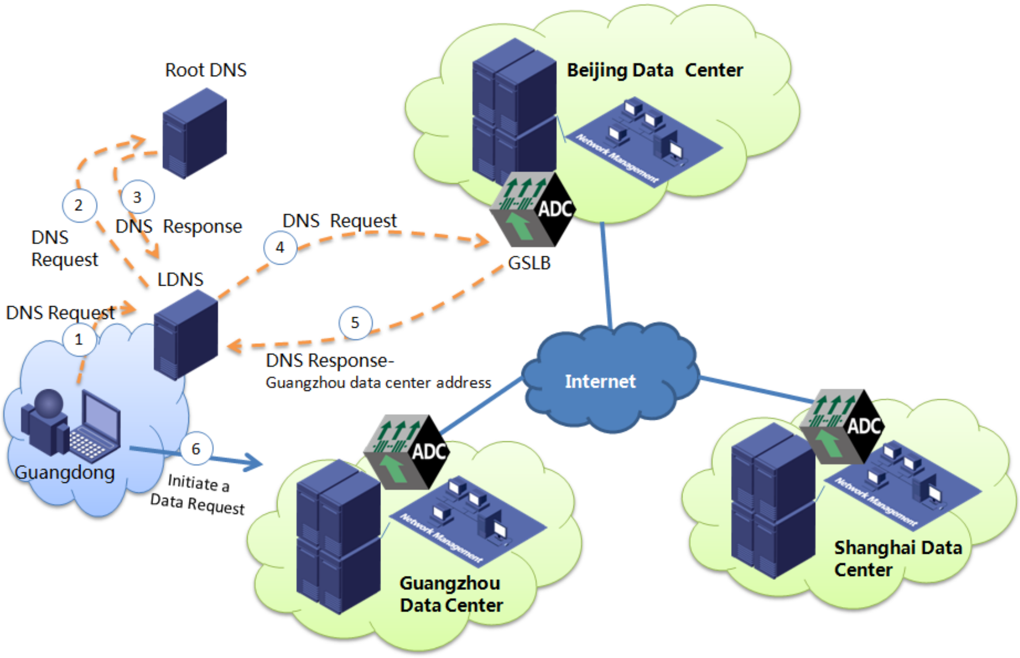


At the same time, Venusense ADC also provides professional anti-DoS, Firewall and Web application firewall functions, thus further enhancing the security and reliability of applications.

2.1.2 Global Server Load Balancing

Global server load balancing - GSLB (Global Server Load Balancing) is a kind of solution that is used to immunize the servers from the access interruption and enhance the response capability of overall service system by deploying multiple data centers around the world. By deploying Venusense ADC, load balancing across multiple data centers can be realized, and redundancy and disaster recovery between data centers also can be performed. Based on intelligent algorithms, Venusense ADC can guide user’s access to data center that is closest to the user and has the minimum delay, thus achievingthe scalability of entire service system and also effectively improving the user experience.

VenusenseADC has provided built-in intelligent DNS systems and IP location information database. The companies can use ADC device as domain authorized publisher, configure multiple IP addresses for data centers to match with DNSA record. When receiving user’s DNS request, it will return the most optimum data center address corresponding to this domain by determining the user’s geographical and combing dynamic detection algorithm or static policy.

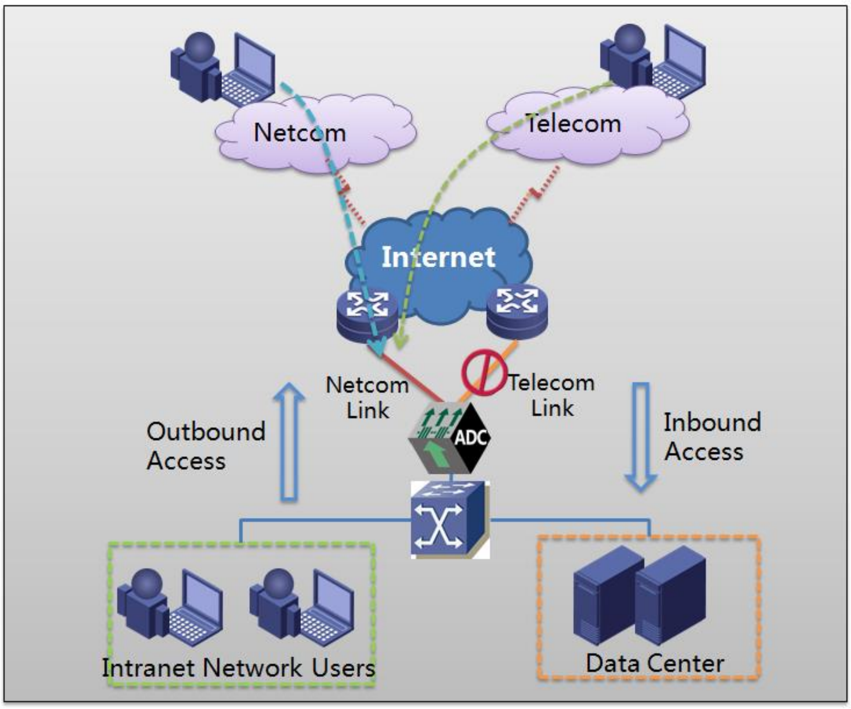


In addition to dynamic intelligent parse mode, Venusense ADC also provides multiple global server load balancing solutions such as static proximity policy, HTTP redirection and IP AnyCast technology. We can provide the most flexible options to the companies, and also have the ability to realize network compatible with other global server load balancing product.

2.1.3 Link load balancing

According to China's ISP access situation, enterprises often choose to rent a number of ISPs' line to realize Internet's link backup and bandwidth overlay. Venusense ADC can dynamically monitor the link status in real time, offers a variety of static and dynamic traffic balancing mode, as well as effectively enhance the efficiency and overall performance of the multi-link access.

When the company deploys the external service application server, Venusense ADC can perform intelligent DNS resolution according to the user's ISP network, geographical distance, or the bandwidth quality of current link to help users choose the best link to access, thus effectively avoiding bandwidth bottlenecks and latency increasing and other problems caused by inter-ISP access, and providing the best user experience.



- **Outbound address access**

  When internal network users initiate external connection request, Venusense ADC products provide a variety of static and dynamic link balancing algorithms to select the most appropriate allocation of link traffic. Static algorithms include: polling, rate and weighting etc. Dynamic algorithms include: Minimum connection, minimum flow rate, minimum delay etc.

  VenusenseADC enables ISP route selection, namely select the matching ISP's link export according to user request addresses, thus avoiding the low efficiency when making inter-ISP access.

- **Inbound access**

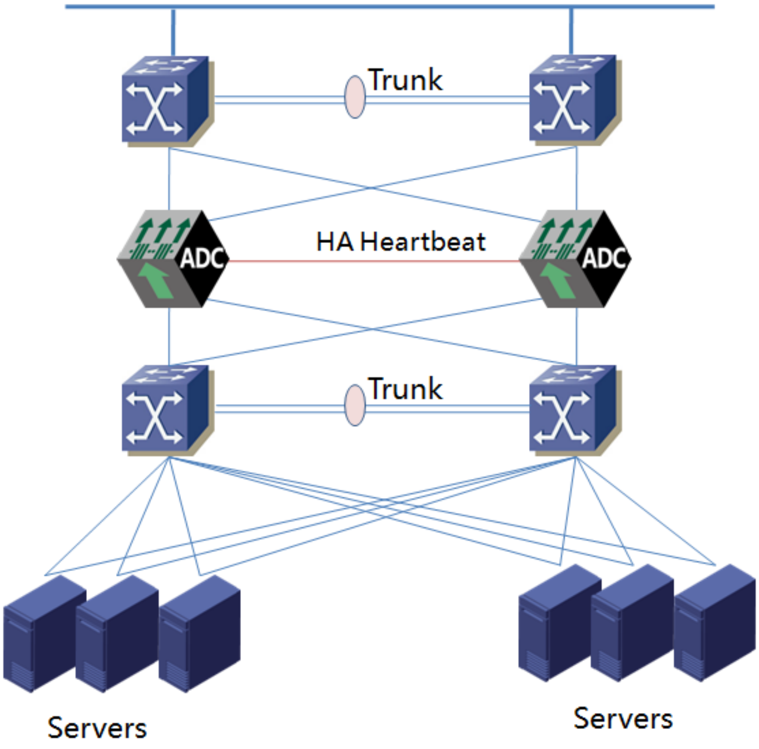
  When the enterprise internally provides external service systems, like ERP systems, mail system or other online service transaction system, the VenusenseADC can be used as authoritative domain name publisher, and bind multiple ISPs' link access IP addresses to the same domain A record. Therefore, after combining ISP's IP location information libraries and static configuration policies, ADC can intelligently handle external user's DNS requests and return the best link access address.

2.2 Deployment mode

Venusense ADC supports serial access, parallel access, L3 access, transparent mode access, DSR mode and other access modes. The enterprise can choose the most appropriate access mode based on the current network health and service plan.

2.2.1 Serial deployment L3 access

In this model, Venusense ADC will connect to the network in serial access mode, and all traffic will be processed by the ADC device. While under normal circumstances, Virtual Services (VS) on the ADC device can be configured as public network IP address, and the internal server is configured as a private IP address. The typical serial network is structured as follows, Venusense ADC will adopt HA mode and realize dual-link cross-connection with the upper and lower switches, which features clear network architecture, and strong redundancy and reliability.

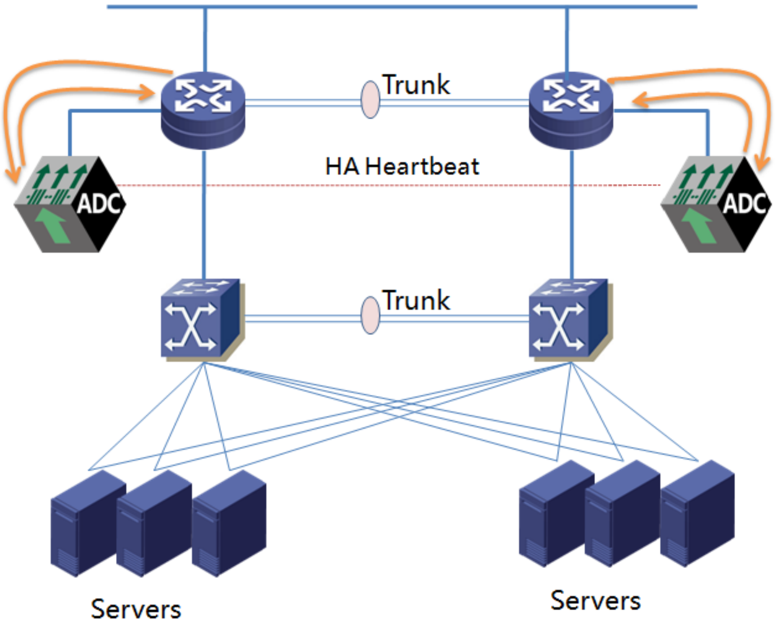


2.2.2 Serial transparent deployment

With this deployment mode, it can make virtual service address (VS) of ADC device and the server locate in a network segment without changing the existing structure of the user's IP address. The working mode of serial access transparent deployment is similar to the normal serial access.

2.2.3 Bypass deployment L3 access

With ADC product deployed in a bypass mode, it can easily and quickly deploy ADC into the network without changing the existing network structure changes, while the ADC working mode is relatively similar to the serial deployment, but the ADC’s Virtual service is configured to public address, and Internal server is configured to private IP address. ADC devices and servers seperately belong to different VLAN of the switch. When the ADC distributes the data to the server, the source IP address will be translated, The message source IP address sent to the server will be changed to the IP address of the ADC device itself to ensure that traffic returned by the server will also pass through Venusense ADC.

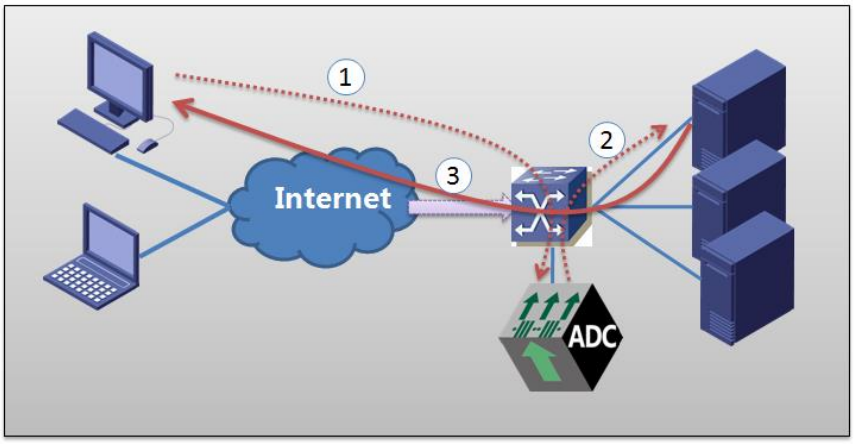


2.2.4 Bypass deployment transparent access

The work mode is similar to the L3 access, except that the Venusense ADC and the server allocates to a VLAN, Venusense's VS addresses and servers are configured to the same IP network segment, which can direct the gateway address of the server to the ADC device's IP address in this case, While the ADC device uses client real address to establish connection with the server, rather than NAT source address translation the NAT source address.

2.2.5 Triangulation of bypass deployment

Triangulation, also known as Direct Server Return (DSR) mode, is a special case ofbypass deployment. In this mode, only inbound traffic will be transferred into the ADC device. The ADC device will distribute the traffic according to the pre-definedload balancing strategy, while the server return traffic will not pass through ADC devices. Due to the typical asymmetry of the Internet traffic(namely the traffic in the request directionis relatively small, and the vast majority of the traffic will concentrate in the direction of the server response) the ADC device will process only the traffic in the traffic request, while the traffic returned by the server will be directly returned to the client rather than passing through the ADC device, thus greatly enhancing the processing capabilities of the ADC.

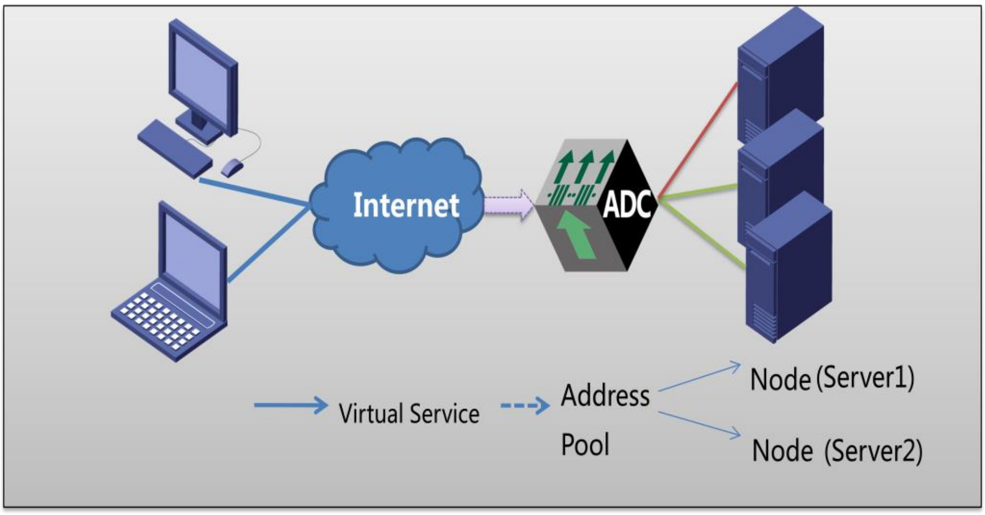


Triangulation mode does not support some functions that need to modify the server return such as Cookie-based session persistence, response rewrite and seven functions. It cannot achieve caching and content compression functions.

Chapter 3 Server Load Balancing

3.1 Server Load Balancing

ADC can create one or more virtual service VS (IP: Port) on the device, to map the internal server pool to provide one or more external applications. Internal servers will be added to the address pool when the external traffic accesses VS. The ADC will select an available server from the address pool to serve as application server according to the pre-defined load balancing algorithm. ADC will simultaneously perform health check for each server node in real time. If an ADC failed to provide service due to its failure, the server will be removed from the available list in the pool and the traffic will not be distributed.



3.1.1 Load balancing algorithm

Venusense ADC supports comprehensive load balancing policy, namely it can realize dynamic algorithm load balancing based on the preconfigured static algorithm and current operating state.

Static algorithms include:

- Round Robin –distribute the traffic to each server in order.
- Ratio) - Specify a weight value for each server based on the performance of the server and distribute the traffic to the server according to the distribution ratio.
- Priority - When multiple pools of servers are used, it will specify a priority to each pool of servers and distribute the traffic to high-priority server pool by default, and select the backup server pool if the sets of server fail.

Dynamic algorithms include:

- Least Connection - ADC will first distribute the traffic to the server with the fewest connections.
- Fastest - ADC will select a server with fastest response to distribute the traffic by comparing the delay of server return packet.
- SNMP monitoring – the device will read the real time operational status of the server via SNMP client, including CPU, memory and I/O information, and configure the threshold for each type of real time information, and will not assign the message to this server if the value exceeds this threshold value.
- Observation mode (Observed) - Select servers for the new request based on the minimum connection and fastest mode, as well as the best balance.

Server smooth access and exit:

When a new server accesses or the server is restarted, ADC system can gradually allocate the traffic to the new access server to avoid excessive system resource consumption or slow application response arising from failing to load some server's processes, thus achieving access to the server smoothly. Administrators can also manually operate to let a server exit the traffic sharing. Meanwhile, the ADC system will no longer distribute new traffic to the server, but the server's existing connections will remain until the connection ends.

3.1.2 Health check policy

Health Check refers to the periodical real time inspection for the operational status of the server. Once a failure server is found, the server will be removed from the traffic sharing. Venusense ADC provides a wealth of health check policies, and can realize very flexible load balancing policy if it combines with the load balancing algorithm.

- TCP SYN-sends TCP SYN packet to the destination server. If the correct response is received, it means that the server is working properly.
- Ping- sends ICMP request packet to the destination server. If the correct response is received, it means that the server is working properly.
- HTTP/HTTPS Get - sends HTTP or HTTPS protocol Get message to the destination system and request a dedicated URL. If the correct response is received, it means that the server is working properly.

3.1.3 Session persistence policy

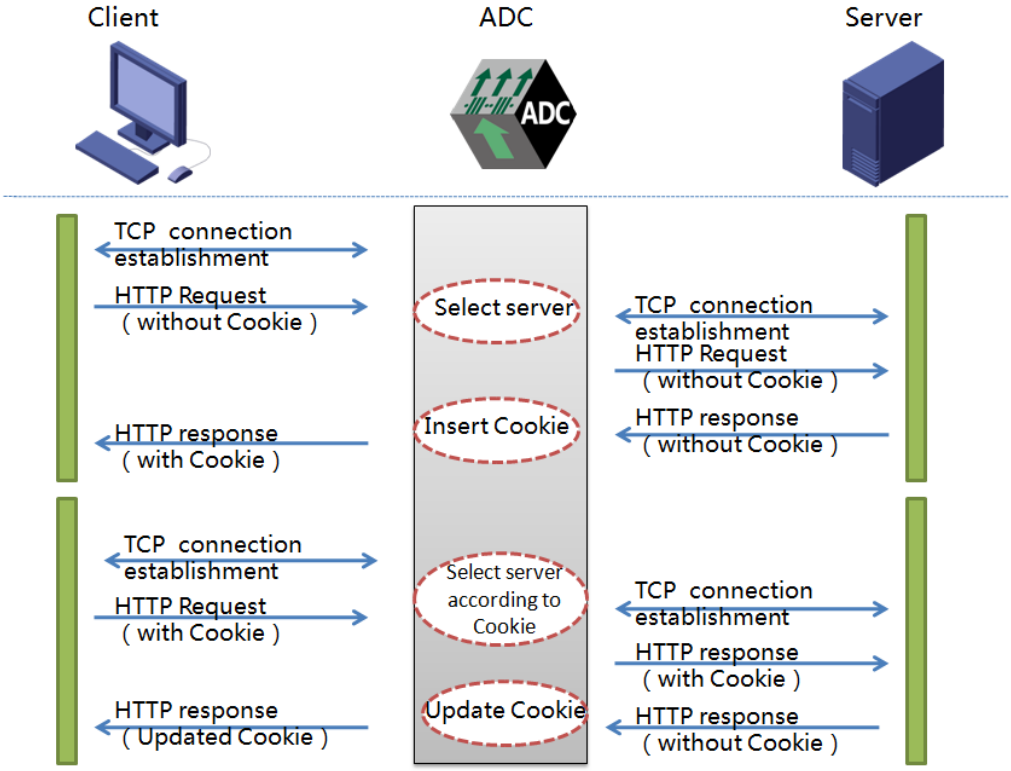
The Session persistence means that the subsequent related request packet will assign to the same server to protect the continuity of service if the traffic is distributed to a server according to the load balancing policy. For example, numerous e-commerce applications or systems that require user authentication, several times of data exchange between user and server shall be performed to complete a transaction or authentication process, and the exchange packets of this process shall be allocated to the same server.

Venusense ADC supports six categories and eight kinds of modes to maintain the session, namely the session can be maintained based on the source IP address, Server ID and other static information, and advanced Session persistence mode can also be realized through inserting and rewrite Cookie. Each session’s retention efficiency, granularity and application scenarios are different, and the application server's configuration requirements are not the same. Session persistence based on source IP address only need to process L4 information, so it has the highest efficiency and does not require the server to make any configuration, but its granularity is relatively large. If the client has widespread NAT translation, or some IP service segments have larger request, this traffic will be kept and sent to a fixed server, thus causing the imbalance of traffic load.

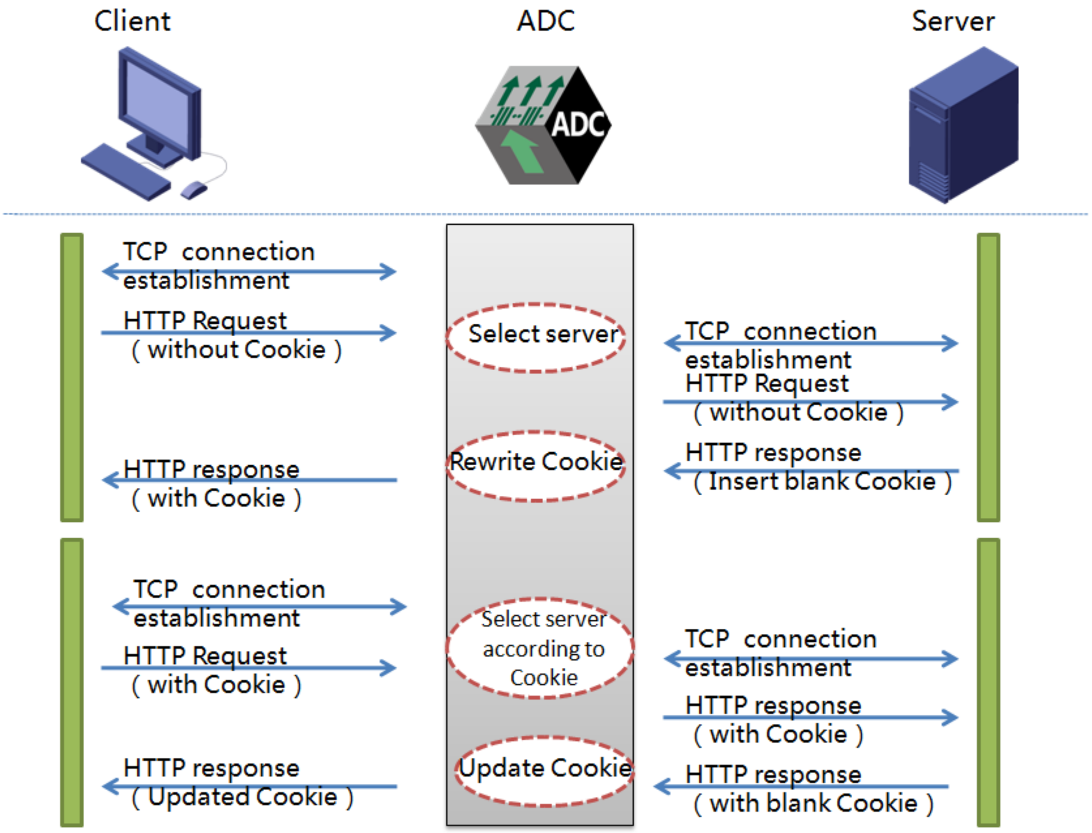
Cookie-based insert, Cookie rewrite, Server ID and other seven information Session persistence mode will make the retention strategy and browser information interrelate, and let the traffic distribution become fine-grained and more balanced. While the working efficiency for the Session persistence mode of seven-layer's information is not efficient as the source IP session. In addition to Cookie insert mode, other mode such as Server ID, sessionid, Cookiere writing etc. generally require the application to make the appropriate configuration.

Venusense ADC supported Session persistence modes:

- Based- source IP address - related packets from the same source IP address will be distributed to the same server.
- Based-Server ID- record the Server ID information returned by the server, look Server ID from URL request packets or Cookie information and obtain backend server information after decoding, and then ensure that requests with fixed Server ID information are distributed to the fixed server. Server ID requires to perform manual configuration on the web server. ADC does not require the modification of the data packets
- Based-Session ID- Server ID similar manner, record Session ID information returned from the server, look Session ID from request packets URL or Cookie information, and obtain backend server information after decoding, and then ensure that requests with fixed Session ID information are distributed to the fixed server. Session ID is automatically generated by server, ADC product does not require the modification of the data packets.
- Cookie-based insertion - The mode enables to modify the packets returned from server, insert fixed Cookie information and then return to the client browser. The client subsequent request in the client contains Cookie information, ADC will send to specified server based on this information. Such retention mode enables to modify the length of the packet, which can be realized without modifying any configuration of application server.


Client
ADC
Server
TCP connection establishment
HTTP Request (without Cookie)
Select server
TCP connection establishment
HTTP Request (without Cookie)
Insert Cookie
HTTP response (without Cookie)
HTTP response (with Cookie)
TCP connection establishment
HTTP Request (with Cookie)
Select server according to Cookie
TCP connection establishment
HTTP response (with Cookie)
Update Cookie
HTTP response (without Cookie)
HTTP response (Updated Cookie)

- Cookie-based rewrite - after the server receives HTTP request, the response packet will be added a blank Cookie to return to the ADC device, and ADC will write the session persistence value in this blank Cookie, and then return to the client. The subsequent procedures are similar to the insert Cookie. The subsequent request packets of the client will contain the rewritten Cookie, ADC will select the specified server based on this information. The difference between Cookie rewrite and Cookie insert is: the requirement for modifying the length of the packet in Cookie rewrite is not mandatory and the efficiency will be higher.


Client
ADC
Server
TCP connection establishment
HTTP Request (without Cookie)
Select server
TCP connection establishment
HTTP Request (without Cookie)
HTTP response (Insert blank Cookie)
Rewrite Cookie
HTTP response (with Cookie)
TCP connection establishment
HTTP Request (with Cookie)
Select server according to Cookie
TCP connection establishment
HTTP response (with Cookie)
HTTP response (with blank Cookie)
Update Cookie
HTTP response (Updated Cookie)

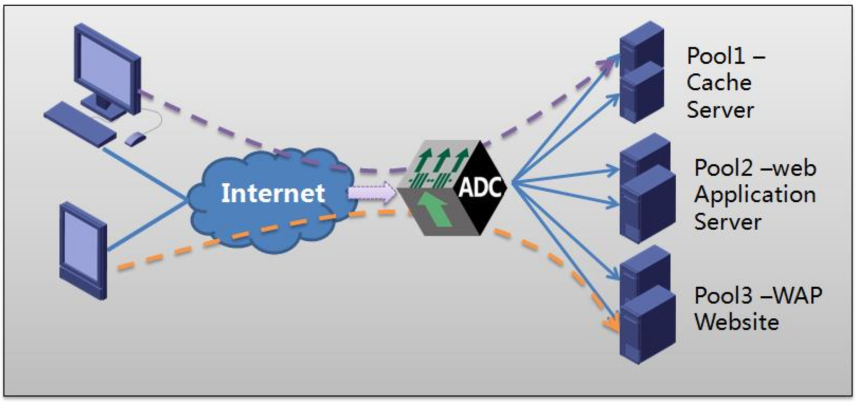
- Based-customized header - application service itself defines a URL Header information, and hope ADC to perform load balancing. In this way, ADC device will record Header information returned by the server, and perform matching operation in the client's subsequent request, and will forward to the specified server if the match succeeds. This allows application service program to customize their session persistence policies.
- Based-SSL Session ID - when the first request arrives, backend server will be assigned according to the load balancing algorithm. In the server's response, it will remove the SSL Session ID according to the SSL protocol, and will save the backend server information assigned by the SSL Session ID and the configured timeout into a table. When subsequent requests arrive, ADC will find backend server information in the table according to SSL Session ID in the request. If the time is found within the timeout period, it will remove the backend server, and update the timeout, and then send the request to coresponding server.

3.2 L7 content exchange

L4 exchange realizes traffic distribution mainly relying on IP and TCP/UDP layer's information, However, the complexity of the application and continuous improvement of user experience demands, sometimes it will require to return different presentation content for different types of users, such as:

- Distribute mobile user's mobile phone/pad browser requests to the specially optimized and destinated server.
- Distribute the request for pictures, documents, videos and other static content to cache servers.
- Return appropriate page to users in different locales based on browser's own language settings.
- Realize read and write separation according to the HTTP request mode, HTTP read (get) requests will be distributed to the cache server, HTTP write (post) requests will be distributed to the server that is responsible for handling dynamic content.

Venusense ADC's L7 content exchange can identify the content of user requests packet, such as URL information, application data types, Cookie information, browser type, HTTP mode etc. and will assign the traffic to the corresponding application server.



Venusense ADC will identify service class via http-class, while the http-class will define according to the host addresses, URI path, header information and Cookie. Each http-class can associate with a server address pool, and then refer one or more of http-classes in the virtual service (VS) configuration. When a client accesses to virtual services, ADC device will perform http-class matches, the matching request will be assigned to the corresponding address pool.

3.3 Application security protection

3.3.1 Status firewall and traffic management

Venusense ADC provides professional-level security firewall features, which can flexibly configure enabling or disabling access policies for each virtual services (VS). Unlike switch or router's ACL, Venusense ADC system will filter the firewall based on the status packet, which can effectively protect the access security of each application service.

Venusense ADC's firewall also includes traffic management features, which can limit the maximum bandwidth and minimum bandwidth for data streams that access to certain type of VS. When there are multiple application systems running in the system simultaneously, it can effectively implement bandwidth restriction and bandwidth reservation for each application in this way.

3.3.2 Anti-L4 DDoS attacks

With the intelligent algorithms, Venusense ADC supports most of DDoS attacks, including TCP SYN Flood attacks, TCP blended attacks, UDP blended attacks and ICMP attacks. With intelligent algorithms, it effectively identifies attacks and normal traffic to ensure the normal operation of the service.

3.3.3 Anti- HTTP - DDoS attacks

Compared to L4 DDoS attacks, L7 attacks based on HTTP are more difficult to prevent, as conventional technology is difficult to distinguish between the normal traffic and attack traffic. One of the attacks is called the CC attack mainly using proxy servers to generate numerous HTTP requests and quickly consume web application server resources, which is the typical attack.

After you enable anti-CC attack, Venusense ADC will dynamically insert a dedicated Cookie in the HTTP response returned by the server. When accessing the browser normally, the Cookie will be carried back and be treated as "chicken's" proxy server and will not be recognizable, and will not be carried in the next request, so that the Venusense can distinguish normal traffic and attack traffic, and the attack traffic will be directly discarded.

3.3.4 Server connection management

Venusense ADC will set the maximum number of concurrent connections threshold for each application, which can not only protect the application server resources, but also can prevent certain applications from consuming too much system resources, resulting in other applications can not perform normal service.

3.3.5 Information leakage prevention

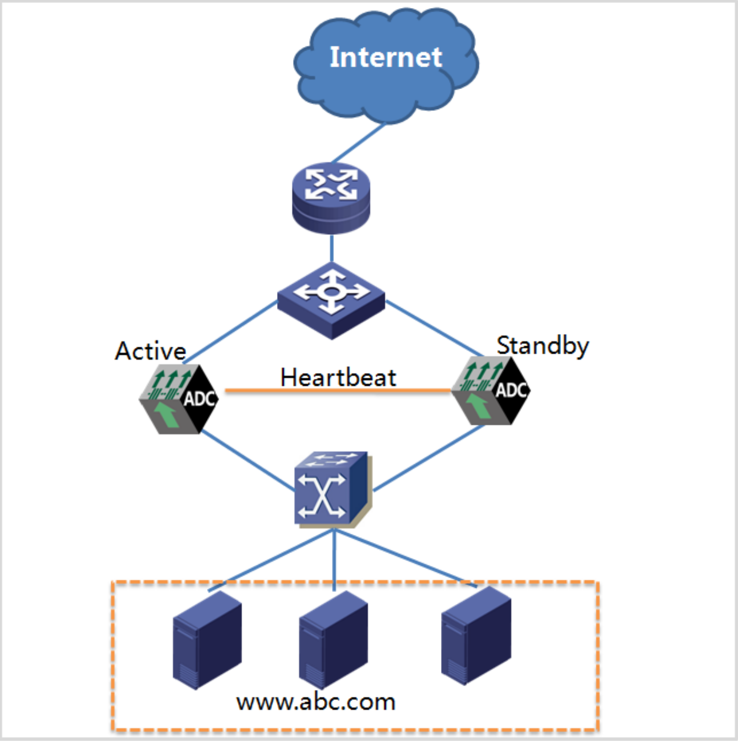
- HTTP header information hiding: Venusense ADC can erase specified information in the response header from the server, such as web server name and version number. It can also modify the real link directory in server to hide critical information. The allowed header can also be defined in the Response, whereas the other header information that does not satisfy the requirement will be deleted from the packet.
- Cookie Encryption: the plaintext Cookie information in the server response packets will be encrypted, and then sent to the client to prevent certain critical information(such as user accounts, online banking data) that is stored with Cookie from intercepting during transmission. In response to the client process, the Cookie information will be decrypted and sent to the application server, which facilitates unified management of application system and reduces the pressure of application system.

3.4 HA high availability and device cluster

3.4.1 HA high availability

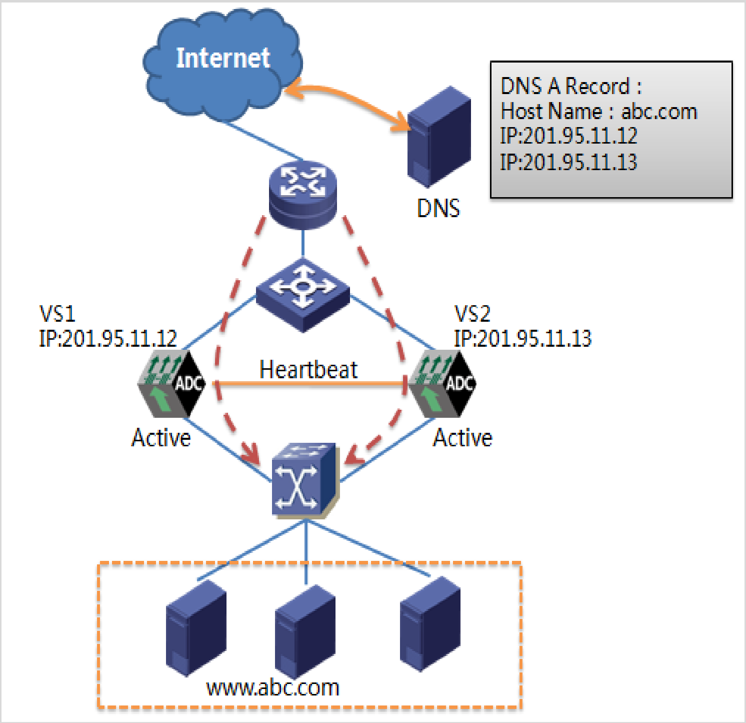
Venusense ADC supports Dual-system Active-Standby and Active-Active modes of operation. During running, HA will monitor operational status of end devices via dedicated "heartbeat" in real time. When the heartbeat monitor fails or other trigger switch condition occurs, the traffic on the failure device will be taken over to ensure that the applications are running uninterruptedly. Venusense ADC’s HA module supports configuration synchronization and connection information synchronization.

- **Active-Standby mode:** one of ADC device is active, another one is in the Standby state, all traffic will be forwarded by the device in the active state. The standby device in standby state will monitor the operating state of master device via "heartbeat" in real time. When the normal heartbeat packet cannot be monitored, the standby device will switch to Active status, and will update ARP cache of upstream and downstream device via free ARP to achieve the takeover of traffic.



In addition to the backup device can initiatively take over when the abnormal heartbeat is found in the master, the master device in active can also initiatively exit from the Active state, according to some other trigger condition. Triggering conditions include: Whether the predefined remote IP address with port status monitoring is reachable..

- **Active-Activemode:** two ADC devices simultaneously are in Active state and will forward the traffic together. When different VS (virtual services) are running on two ADC devices, and one of them fails, all Virtual Service traffic running on the device will be taken over by another device. Under normal circumstances, the master-master mode can be used with DNS load balancing to work together.

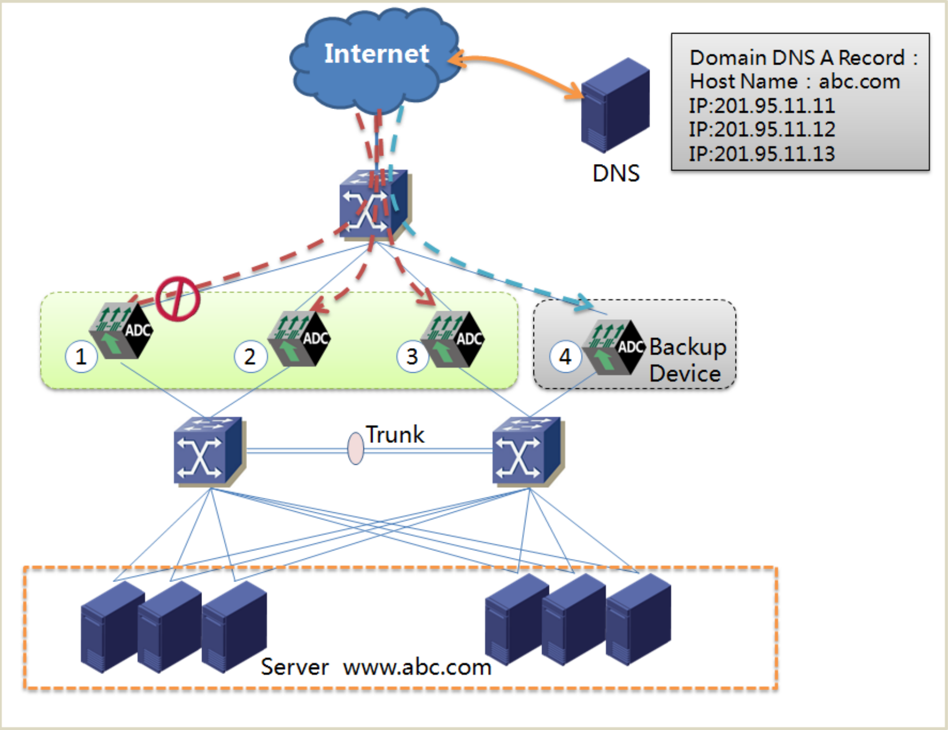


When users request to access www.abc.com, it shall first initiate DNS request, host name corresponding to DNS server is configured as abc.com. This record will configure two IP addresses corresponding to VS, and DNS server will use polling mode to return one of IPs to realize load balancing.

3.4.2 Device cluster

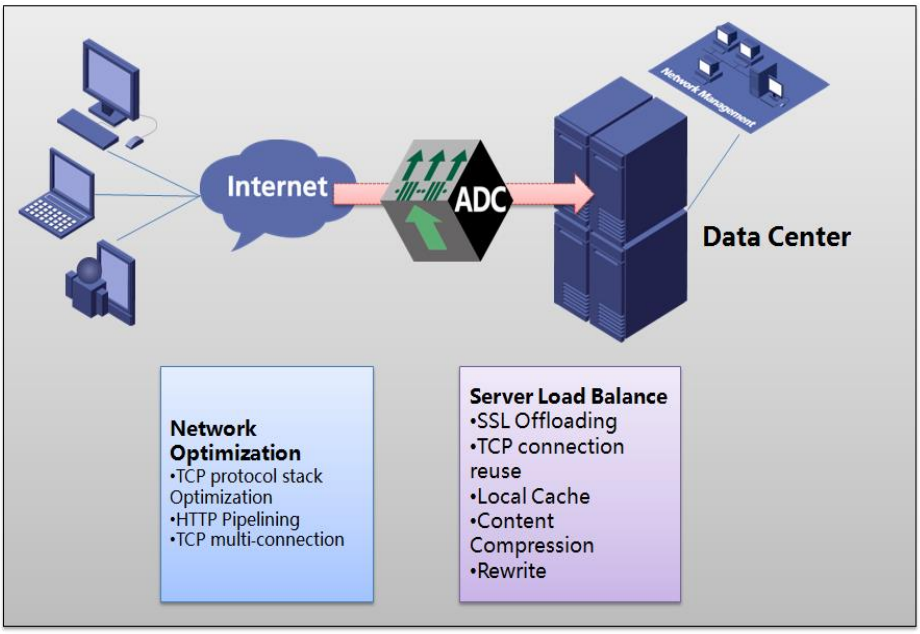
Venusense ADC can realize clustered deployment up to 32 ADC devices, and multiple ADC devices can share the traffic processing and redundancy. ADC devices in the cluster can be different models and with different processing capabilities, so for the upgrading devices, they can also be reconnected to the cluster. Through the cluster deployment, enterprises can maximize service resilience and maximize device investment income.

The full operating mode can be adopted when deploying the cluster, i.e. all devices can participate in forwarding traffic and backup for each other. If one of them fails, traffic will be switched to the rest devices; N + 1 mode can also be used, i.e. a device can be used as backup for other N sets of devices. Under normal circumstances, it will not back up the device and will not forward traffic, and will take over the traffic when one device fails. N + 1 mode can prevent traffic peak in another device when a device is malfunction.



Chapter 4 Application Optimization and Acceleration

Venusense ADC can offload the task that originally requires high consumption of server computing capacity, duplicated computing to high-performance hardware platform, thus letting server computing resources focus on their service processing, so as to improve the efficiency of overall application system. Venusense ADC can simultaneously optimize and accelerate TCP and HTTP protocols, and maximum reduce the network congestion and packet loss, thus improving user experience for mobile user and other narrowband(mobile) user.



Venusense ADC has provided built-in common enterprise application templates, such as BEA Web logic, Microsoft IIS, Outlook Web Access, Radius, ERP software, which is a complete set solution that is repeatedly tested and verified by our company's application optimization experts. Administrators can effectively optimize and accelerate the application system with these templates without deep understanding of the application.

4.1 SSL hardware acceleration and offload

Venusense ADC will realize SSL protocol's acceleration and offloading via professional-level high-performance hardware acceleration chip. The transmission between ADC and server will be in plaintext inside the data center, thus greatly improving service processes ability of server. The enterprises can adopt SSL protocol for all applications and realize high security, which will not bring any performance bottlenecks to service.

4.2 Local RAM Cache

Local RAM Cache will be used to store some static files (pictures, documents, video files etc.) in the server by opening up some dedicated memory (RAM)space on the ADC device. After enabling local Cache, the client request will first find in the local Cache, and return directly to the client after having found. If it fails to find, it will send a request to the server, while the content returned from server will perform local Cache. Venusense ADC supports respective Cache space and parameter settings for different applications, which can release server from repetitive processing and improve overall performance.

4.3 Content compression

After compressing response data of HTTP, it can effectively enhance bandwidth utilization and reduce download time. With the high performance compression technology provided by Venusense ADC, it can maximize bandwidth utilization by 80%, and improve application performance 4 times. It can also use compression algorithm between the client browser and the ADC device to realize secure transmission.

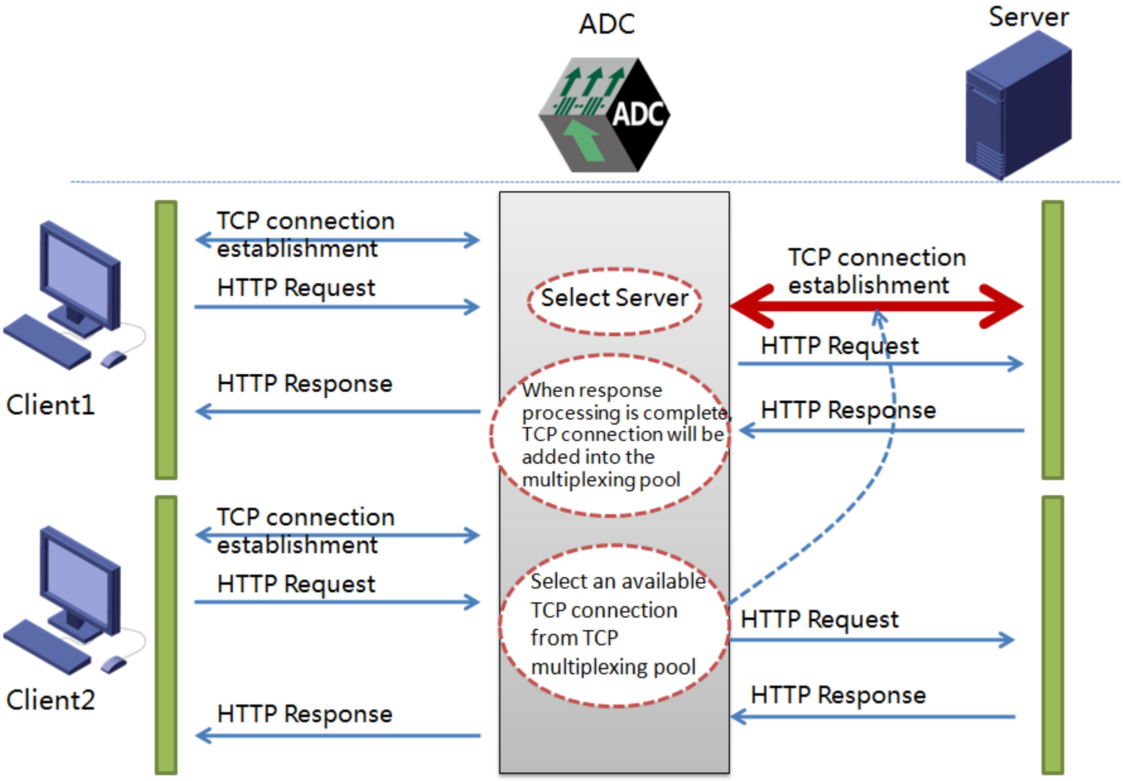
Venusense ADC will move the compression process originally completed by the server to high-performance hardware platform, which avoid implementing repeated compression process in each server, thus achieving server offloading effect.

Venusense ADC browser supports the most commonly used GZIP and DEFLATE compression algorithm, and provides L7-based fine compression control strategies, including URI, Content Type etc. Administrators can define to compress some document types such as "\.txt", "\.doc", "\.html" and static pages data compression, and also exclude some unnecessary operations with ineffective compression such as PDF, IMG ,etc.

4.4 TCP connection multiplexing

TCP connection multiplexing enables multiple clients to share one TCP connection to the server to enhance the overall performance of the application server, so that the application server can release from the maintenance for massive TCP connections and continuously implementing TCP connection and disconnection, thus significantly improving the load capacity of single server.

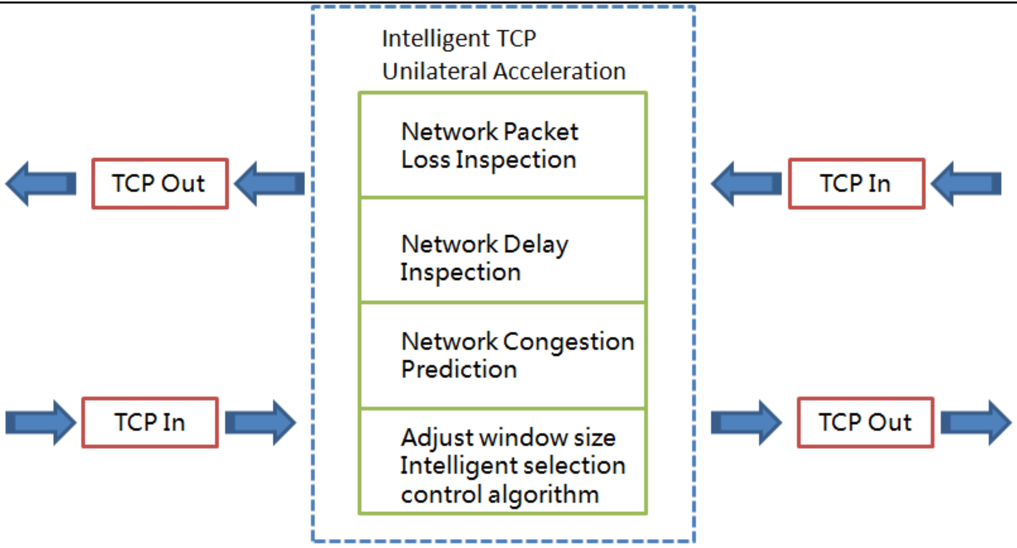
When the Venusense ADC device receives client HTTP request, it will chooses a server to establish connection through load balancing algorithm. If the server receives normal response, ADC device will put this connection into the "connection multiplexing pool". When another new HTTP client initiates a connection request, ADC will select an available connection from the existing connection pool, and exchange data with server rather than re-create a connection to the server. In this way, the server load can be reduced to the original 1/10 to 1/50.


ADC
Server
ADC
Client1
Client2
TCP connection establishment
HTTP Request
HTTP Response
Select Server
TCP connection establishment
HTTP Request
HTTP Response
When response processing is complete, TCP connection will be added into the multiplexing pool
TCP connection establishment
HTTP Request
HTTP Response
Select an available TCP connection from TCP multiplexing pool
HTTP Request
HTTP Response

4.5 TCP unilateral acceleration

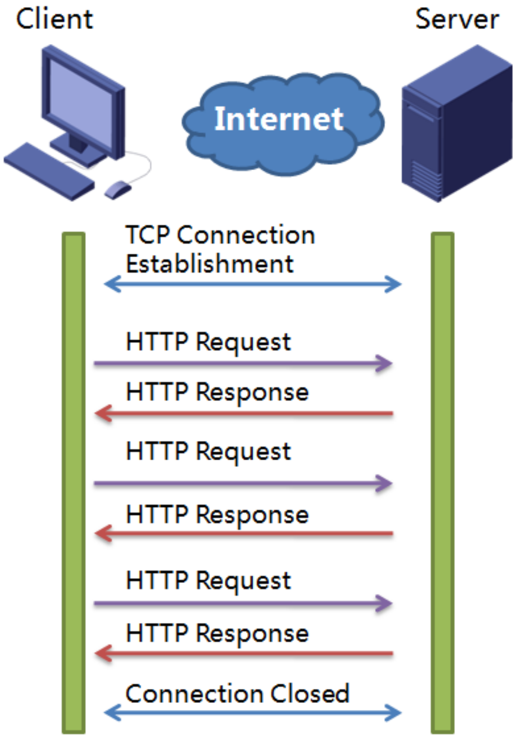
Standard TCP protocol has not fully considered the broad bandwidth and cross-Internet transmissions in design. With the popularity of broad bandwidth, standard TCP protocol have many deficiencies, resulting in relatively low efficiency in the network congestion control and packet loss retransmission mechanism, and basically losing the ability to mediate in real time according to the network environment. Venusense ADC provides unilateral intelligent acceleration functions, which can improve the effect of overall network acceleration via optimizing standard TCP protocol slow start control, congestion avoidance, retransmission and recovery. Meanwhile, Venusense ADC devices can intelligently choose the most efficient algorithm based on current network status and carrying protocol type.

Venusense ADC’s unilateral acceleration is transparent to both the client and server, ADC and server will run based on the standard TCP protocol, then ADC and the client will accelerate unilaterally, while the client does not required to make any modifications.


Intelligent TCP
Unilateral Acceleration
Network Packet Loss Inspection
Network Delay Inspection
Network Congestion Prediction
Adjust window size Intelligent selection control algorithm
TCP Out
TCP In
TCP In
TCP Out

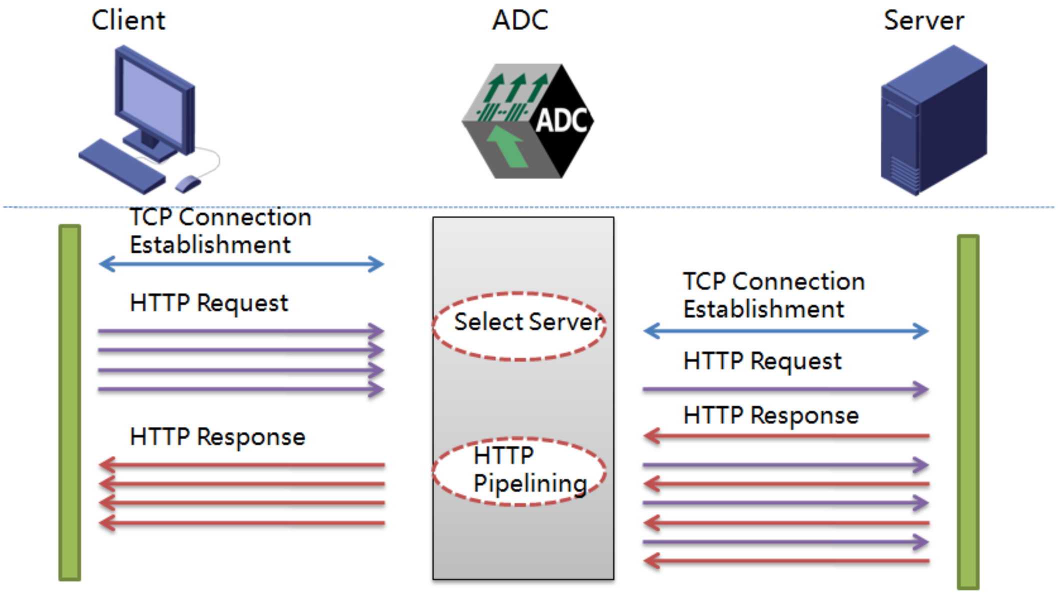
4.6 HTTP Pipeline

Traditional HTTP protocol will process next request until the current request is issued and response data is completely received, which will result in relatively low efficiency of data exchange in the high latency network environment. HTTP pipeline will submit multiple HTTP requests synchronously without waiting for request to respond in order.



HTTP processes without Pipeline

Venusense ADC can intelligently negotiate with the browser that supports Pipelining client to enable, while the ADC and the server are still using the normal HTTP protocol process.



4.7 Narrowband User Application Acceleration

When a client uses narrowband such as ADSL line dial-up or smart phones, tablet terminal to connect to the Internet, the response of the application server will result in packet congestion and loss due to lack of bandwidth on the client side and poor quality of bandwidth, meanwhile,, the client initiates a retransmission request, if a large number of narrowband users simultaneously access, application server will bear huge pressure and reduce its efficiency arising from repeatedly processing these retransmission requests.

Venusense ADC uses TCP data caching technology, which can automatically detect the client’s processing capacity against the server response. For narrowband users, ADC will open up specified buffer space to store the returned data from task server, in this way, the whole data exchange procedure will be implemented in an appropriate speed, thus avoiding a large number of retransmissions. As long as the server completes processing user's request, it can be released to handle other work and neglect the retransmission packet loss.

4.8 Rewrite

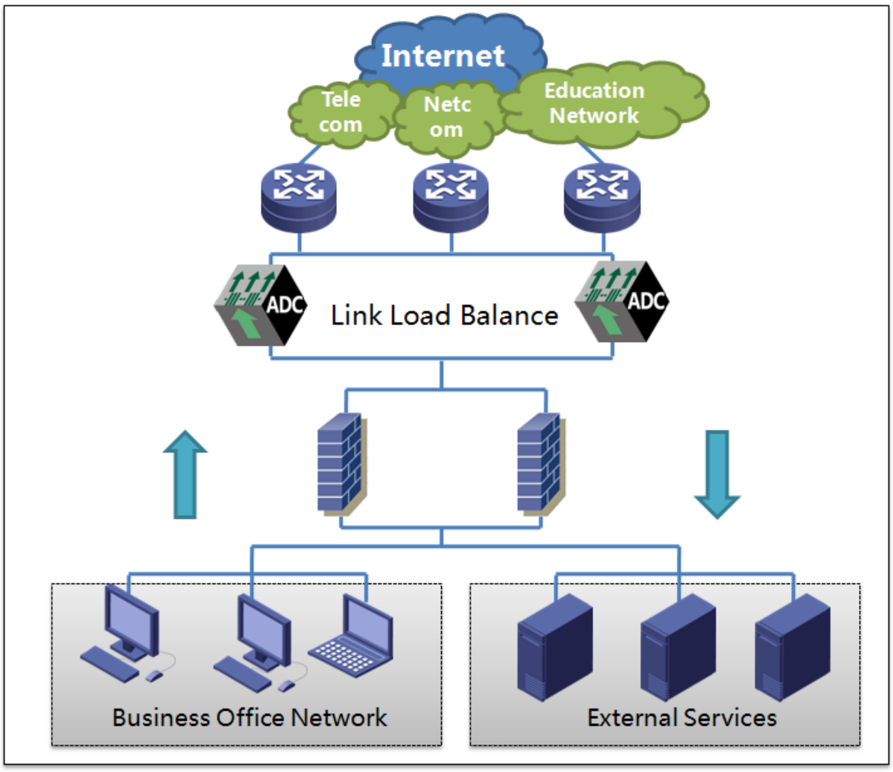
The URI rewriting is feasible during the request and response procedure of the HTTP protocol. However, it can send all http requests that meet all preset conditions to a specified URL during HTTP request procedure, such as user login page or other information prompt page. Or when some server's files do not exist or its directories are changed, and user access through other websites or old links found with search engines, such requests can be directly redirected to the specified error message prompt page, thus reducing the burden on the server.

After the deployment of the Venusense ADC, the enterprises can move the original http protocol applications to the secured https protocol. With the combination of http request redirection and SSL offloading, the enterprise servers and clients can realize seamless migration from http to https without making any changes.

At http response phase, 404 (client syntax errors) returned by the server will be redirected to a specified page.

Chapter 5　Link Load Balancing

With Venusense's ADC product, it can dynamically monitor the real time link status, it offers a variety of static and dynamic traffic balance mode, thus effectively improving the efficiency and overall performance of multilink access. When the enterprises deploy application server for external service, Venusense ADC can perform intelligent DNS resolution according to the user's ISP network, or geographical distance, or the quality of bandwidth of current link, and help users select the best link to access, thus providing the best user experience.

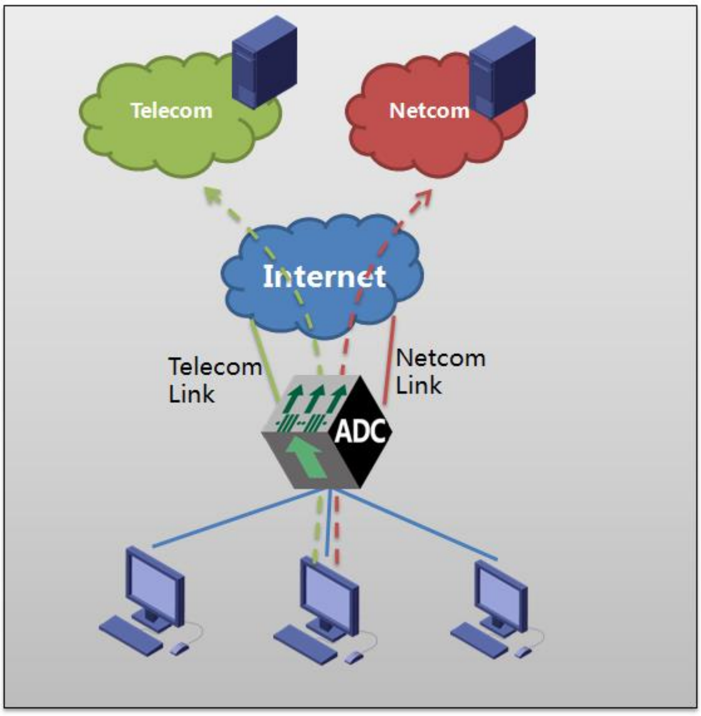


5.1 Outbound traffic load balancing

5.1.1 Load balancing algorithm

Link load balancing algorithm is a kind of program that is used to calculate the traffic distribution across multiple links when network users access the Internet (outbound traffic). Link load balancing algorithm and server load balancing algorithms are identical in some places but there are still some differences, link load balancing algorithms include:

- Round Robin - evenly distribute the traffic to each link in order.
- Ratio - specify a weight according to the bandwidth of each link and distribute traffic to multiple links according to this ratio.
- Priority - specify a priority for each link, distribute the traffic to link with high priority by default, and will select the backup link when the link fails.
- Weight Least Connection - first specify weighted value of bandwidth for each link, so the allocation of number of connections is matched with the set weights. For the new connection, it will select the minimum link within the weight value to allocate traffic.
- Weight Least Traffic - first specify weighted value of bandwidth for each link, so the allocation of traffic is matched with the set weights. For the new traffic, it will select the minimum link within the weight value to allocate traffic.
- ISP Route - provides built-in IP address ISP correspondence table, select the appropriate link based on destination addresses' ISPs that the intranet user access, thus avoiding inter-ISP access.



- Fastest mode - ADC will select the fastest response to allocate traffic by comparing the delay and hop count of data packets returned by the server.
- Standby mode - By default, traffic will be sent to the master link, and will enable backup link when the primary link fails.

5.1.2 Link health check

Venusense ADC link load will monitor and perform health checks the export link in real time, thus being able to discover port down situations. In addition, Venusense ADC supports monitoring the remote address that use multiple protocols including ICMP, TCP SYN, UDP, HTTP Get and other protocols, and can discover the exception and switch the traffic to other available link even if the internal ISP network fails

5.1.3 Outbound traffic session persistence and NAT

The interface traffic session persistence means the some type of subsequent related traffic will be distributed to the same link. Once data stream is allocated to one outbound interface link. Venusense ADC's supporting session persistence mainly relies on the source address session persistence and hash-based session persistence.

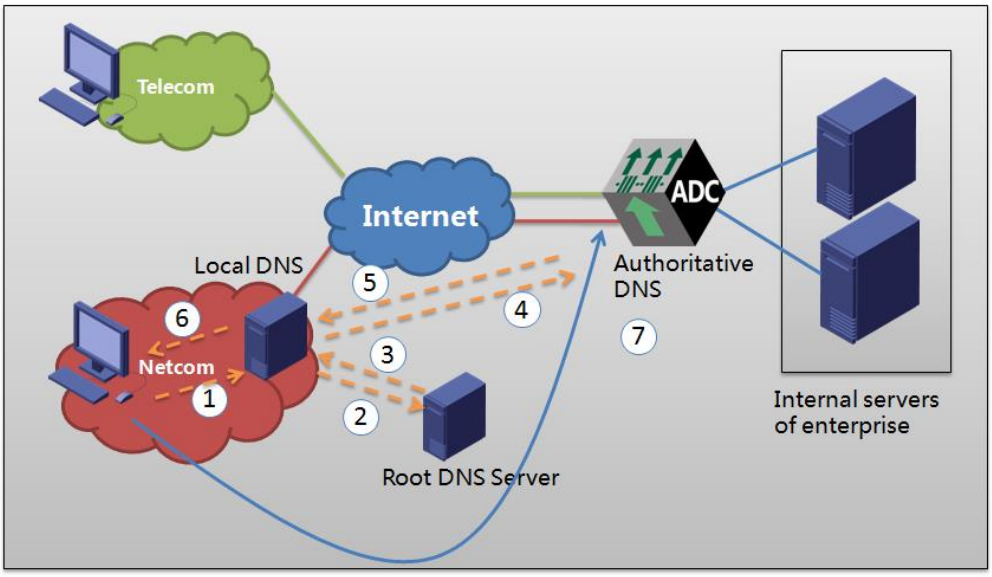
Venusense ADC supports comprehensive NAT translation policy, including source

IP address translation, static address translation and policy-based address translation. Session persistence and NAT address translation can avoid the packets transmitting across multiple ISP to large extent during the communication process between source host and destination server.

5.2 Inbound Traffic Load Balancing

Enterprises usually rent multiple ISP links to establish the service for external users via internal application server, and Venusense ADC will use built-in Intelligent DNS server to bind enterprise's single domain to respective public IP address of each ISP, and regard it as the corporate domain's authoritative publisher. When a client accesses to the application server, it will first make DNS resolution, and then Venusense ADC will return matched IP address according to the ISP network of client, or select user's best communication quality links through dynamic detection technology and return the corresponding IP address. This ensures that each client can access internal application server via the best communication quality link, thus greatly improving the user experience, and enhancing the efficiency of the entire system.

5.2.1 Intelligent DNS resolution process



Assume that the enterprises rent Unicom, Telecom links to provide services to the external network and the enterprises' domain name is www.abc.com, the entire analytical process is shown as follows:

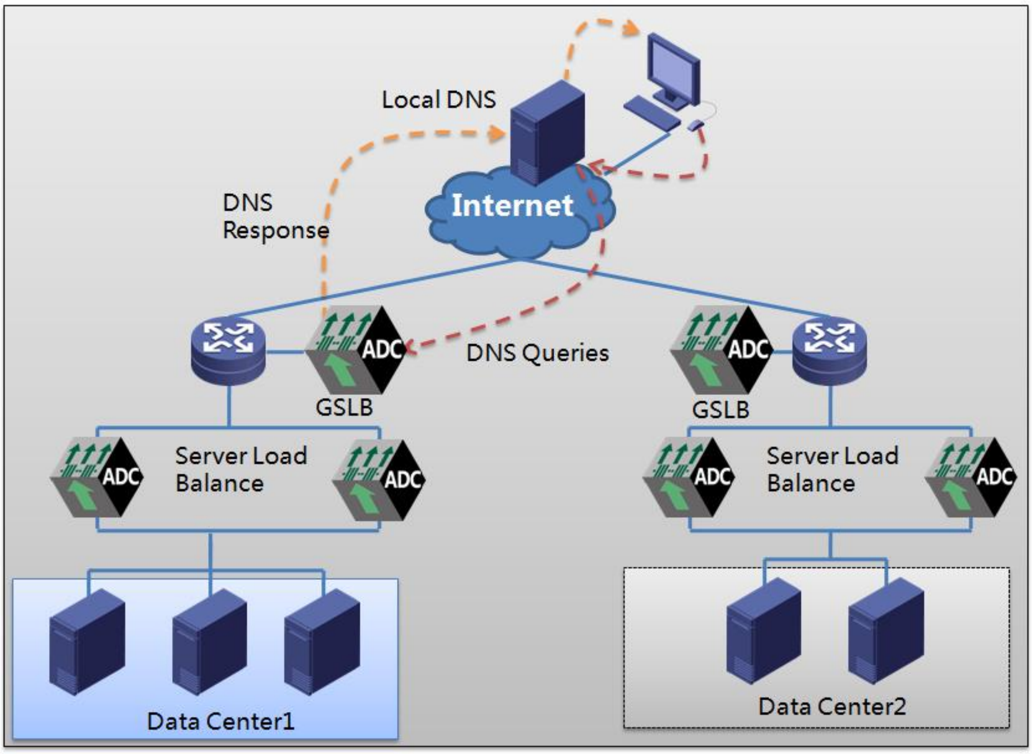
1. The client will send the DNS request with domain name of www.abc.com to local DNS (Local DNS) server

2. LDNS has not kept the domain name’s A record, but will send the request to the server with the maximum matching result, assuming that the maximum matching is the root server.
3. Root servers have not kept the "www.abc.com" A record, but kept the "www.abc.com" NS name server record - "www.abc.com IN NS master.abc.com", and this NS record will be returned to LDNS.
4. LDNS server knows where to find "www.abc.com" A record according to the NS record, and the request will be sent to the intelligent DNS server inside the Venusense ADC device.
5. ADC device will judge which ISP this LDNS IP address is belonging to. In this example, LDNS belongs to China Netcom, so the "www.abc.com" A record returned by Venusense ADC is the public IP address assigned by the Netcom link according to the pre-defined rules.
6. LDNS will eventually send the DNS response that just receives back to the client.
7. The client will access the enterprise’s Netcom link public addresses, and the virtual services (VS) corresponding to Venusense ADC will receive the data and enter into the server load balancing process.

Chapter 6 Global Load Balancing

In order to protect your enterprise from website interruption and improve the performance of applications, the deployment of multiple data centers is an effective approach. But the comprehensive realization of these goals requires your enterprise to monitor the status of infrastructure and applications in an efficient way, and control this distributed infrastructure according to the service demand.

With the global load balancing feature, the safer, smarter respond mode can be provided to user when performing every DNS query. Venusense ADC will assign end-user application requests based on user location, service policies, data center conditions, network conditions and application performance. In this way, the user can fully control the WAN traffic to ensure that the applications distributed across multiple data centers have high availability and maximum performance. So, higher application performance, shorter downtime and more simplified management can be achieved.



6.1 Global Load Balancing Policy

Venusense ADC supports a variety of dynamic and static global load balancing algorithms to provide users with rich options, including:

6.1.1 Dynamic proximity

With the dynamic proximity, the LDNS can be inspected from various global sites, and the response speed for LDNS and each site can be shown based on obtained dynamic parameters, and the following modes are provided depending on the dynamic parameters of reference:

1. **RTT**

With this algorithm, it can dynamically get the response time of LDNS and each data center, and the A records corresponding to IP with smallest RTT value shall be returned to LDNS. RTT stands for Round Trip Time, which record the response time that a website sends probe packets and receives the probe packet. However, the LDNS's RTT detected in different sites' VS in the actual operation reflects the response speed of each VS. The LDNS proximity distribution table will be created dynamically after selecting RTT algorithm. According to this dynamic table, the fastest link will be provided to each user when accessing.

A variety of detection modes can be adopted to determine the LDNS'RTT time, including:

- DNS_Dot: Initiate a test that contain to local DNS, i.e. request root lists from destination, the DNS server with the default configuration usually supports this detection mode.

  DNS_REV: Initiate local IP's PTR requests against local DNS
- UDP: initiate a UDP packet and record response time
- TCP: initiate a TCP packet and record response time
- ICMP: initiate ICMP ping and record response time

2. Hops

Hops refers to hop count from the website to LDNS, which is similar to the RTT. Choose the VS with the minimum hop count from the detected hop count from VS to LDNS. This mode is achieved by traceroute.

3. Global availability

Global availability algorithm is mainly used for disaster backup system. Through health checks, you can determine the health status of each nodeor link. When configuring, IP addresses corresponding to the same domain name will be sorted, and the server that ranks No.1 will provide external service when the system is running normally. Only when the server that ranks No.1 fails to provide external service the server that ranks No.2 will take over the service. If there are multiple lines and nodes, it will perform like this.

Generally, we use the global availability algorithm as an alternative algorithm. If the two above-mentioned modes are not selected, all users will be navigated to the default line.

4. Fall Back IP

Users can configure an IP address, the device will use the A request corresponding to this IP address to respond the LDNS's request if you select this mode. Normally this kind of mode is used as alternative mode. When setting Fall Back IP as the IP of disaster backup data center, this mode will be used only when the preferred mode for dispatching fails, and then LDNS will finally get a response.

5. Poll

The request will be returned to each VS IP sequentially and cyclically. If one of these VSs fails, it will be removed from the cycle queue and will not participate in the next poll until it returned to normal.

6. Weighted Round Robin

Assign a weighted value for each VS, and assigned each user's request to VS based on this weight. Larger weight value means that more requests can be handled, while smaller weight valule means that less request will be processed. When one of these VSs fails, it will be removed from the queue and will not participate in the next user's request distribution until it returned to normal.

7. **Minimum number of connections**

Pass the new connection to those VSs that process minimum connection. When one of these VSs fails, it will be removed from the queue and will not participate in the next user's request distribution until it returned to normal. In order to avoid VS collapse due to overload, it can specify the maximum threshold values of connection for each VS to avoid overload.

8. **Weight least connection**

The algorithm is a superset of minimum connection algorithm, each VS will use its corresponding weight to represent processing performance. The default cost is 1, the system administrator can dynamically set VS's weights. Weighted least-connection scheduling will make VS's established connections correspond to its weight when scheduling new DNS request.

9. **Minimum bandwidth**

Select current VS that utilizes minimum bandwidth. Units kbytes/s.

10. **Minimum traffic**

Select VS that has minimum traffic recently. Units pkts/s.

6.2 Static proximity policy

When selecting static proximity as global load balancing's scheduling, it is equivalent to choose the VS's IP according LDNS's ISP, location and other topology information. Different VS may be located in data center with different geographic or in the same data center that has different ISP link interface.

Under this circumstance, LDNS's ISP and geographical location and VS's ISP, geographical circumstances rarely match exactly, which will require the user to configure static proximity policy. Therefore, it is required to artificially delineate distance relationship between LDNS area and various data centers.

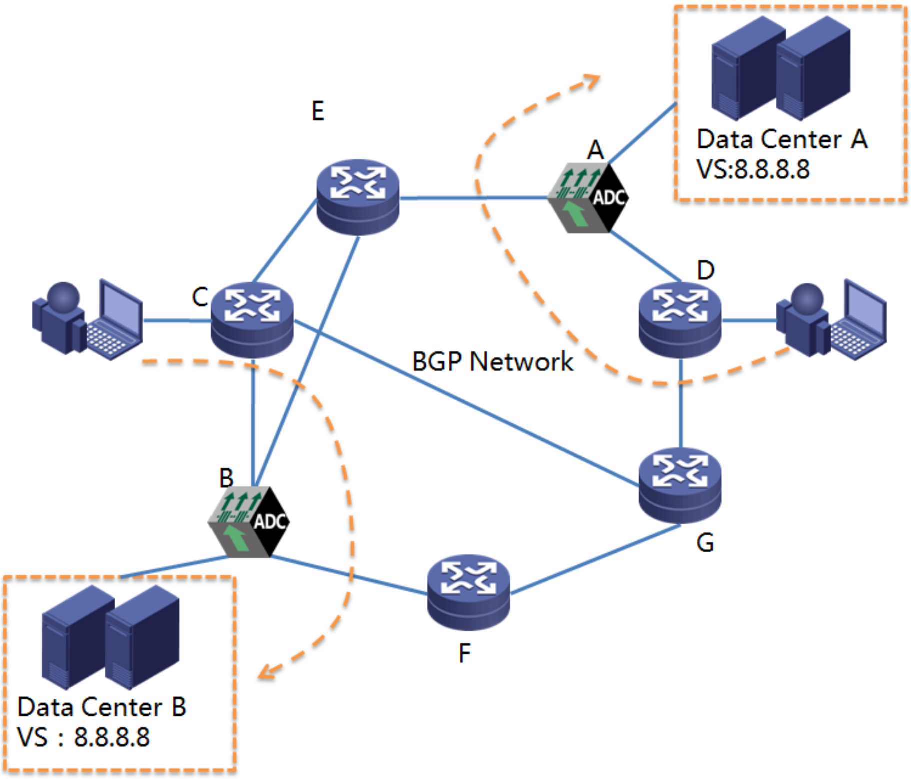
For example: the enterprises deploy two data centers in Beijing and Shenyang and wish data centers in Beijing to cover North China region, while the data center in Shenyang to cover entire northeast region. After combining with the IP address -

geographical information database, it will artificially allocate the LDNS requests from three northeastern provinces to Shenyang data center, and locate the DNS requests from North China to the Beijing data center.

From another angle, the topology information database only provides a mapping relationship between IP and geographical information, and the mapping relationship between geographical information and geographical information shall be provided if it wants to realize load balancing according to the static proximity, and this relationship is called proximity policy.

6.3 IP Anycast technology

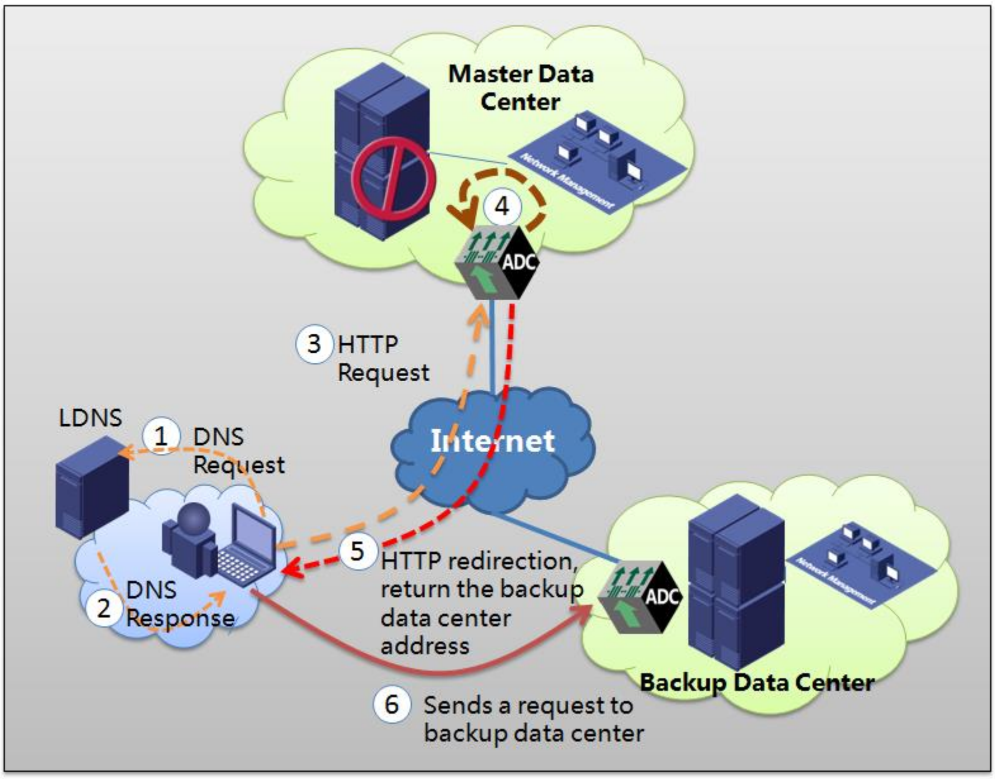
Combined with dynamic route protocol BGP or OSPF, Venusense ADC provides a global load balancing policy (IP Anycast) that is different from the conventional idea. This technology allows the VS deployed in each data center to use the same IP address, and run BGP dynamic route protocol on VenusenseADC, and each ADC device will select the closestdata center according to the route protocols.



IP Anycast technology can also be used as an effective mode to resist DDoS attack. When an attack is launched at any point, it will only affect data center that is located closest to the attack point, and other data center will not be affected and can continue to provide users services.

6.4 Based- HTTP redirection global load balancing

When the user accesses the virtual service(VS) address of application system viaHTTP protocol, the local ADC can use local HTTP protocol's redirection function to re-select an available data center to achieve the requested secondary scheduling if the local ADC fails to schedule. i.e.no available server node can be selected from the address pool or local application system pressure has overload.



With the global load balancing technology based on HTTP redirection (independent from the DNS system, and free of making any modifications to the existing DNS system), it can realize rapid deployment and effective management.

Chapter 7 Application Security Protection

7.1 Application Security Protection

With the rapid development of Internet technology, network application has widely used in the global economy and played an important role in the development of various industries. Over time, application security problems are becoming increasingly serious.

An increasing number of internal and external service applications of enterprises adopt B/S architecture that combines Web and database. Web system is playing an increasingly important role. At the same time, a growing number of Web systems are frequently suffering from the various attacks subjected to security risks, resulting in the leakage of Web system sensitive data and page tampering and even becoming the puppet of spreading Trojans, ultimately will cause damage to more visitors and serious losses.

The Enterprises commonly used network security devices such as firewall, IPS against security threats. However, the traditional network security products mainly work under the network layer, and can identify and control the protocol, address and service port and achieve the purpose of invasion defense, thus effectively protecting service port attack attack. However, facing to the development trend of security threats, it is proven that the firewall is powerless and fails to detect the malicious threats and attacks encapsulated in valid data. Intrusion prevention products are mainly used to prevent network intrusion from comparing features. However, with the deepening and popularity of Web application technology, the speed for discovering Web application vulnerability and launching attack experiences fast,and Web-based vulnerabilities are vulnerable to exploitation andbecome Hackers' prime choice. The attacks against application layer establishment such as Application layer SQL injection, XSS attacks will lead to the traditional intrusion prevention products to ignore such attacks due to the non uniqueness characteristics of these attacks.

7.2 Venusense Application Delivery Solutions

Traditional load product basically does not contain security features or only provides basic network access control. Venusense also provides standard firewall anti-scan, anti-attack capabilities in addition to basic network layer access control, which can completely replace the firewall. As a kind of application delivery product, VenusenseADC will not only perform application distribution, but also support security testing and access control for application layer.

VenusenseADC uses a set of HTTP session rules. These rules cover common Web attacks such as SQL injection and XSS. While more attacks are identified and prevent viacustomizedrules. Web system security problems that cannot be solved with traditional device such as firewall andIPS are solved.

VenusenseADC is committed to providing Web security and application delivery integration solutions to ensure the availability, performance and security of Web applications or network protocols:

- Web Security: Based on years of accumulation in safety testing and sustained investment in security research, it provides fast and comprehensive Web security testing program against SQL/XSS attacks that threat Websecurity.
- Application delivery: relying on the latest intel Sandy Bridge platform and TCP/IP protocol stack optimized for TBOS and focusing on high speed, high availability, it has optimized service resources and improved access experience.

7.3 Major security features

7.3.1 Network layer protection control and attack protection

Venusense provides standard firewall featuresto make access control based on source, destination IP, protocol, port, time and interface information. In addition, it can determine whether the packet is offensive by analyzing the behavioral characteristics of network layer packets, and take measures to protect the network host or network device against attacks.

Currently, common network security threats in the Internet are divided into the following three categories:

DoS attacks

DoS attacks are using a large number of data packets to attack the destination system, which make destination system fail to accept the normal user's request or suspend destination host and work improperly. Major DoS attacks include SYN Flood, Fraggle etc. The difference between DoS attacks and other type of attacks is the attacker will not seek entry to the destination network rather than disrupt the normal working of the destination network and prevent legalusers from accessing network resources.

Scan attack

Scan attacks will use ping scanning (including ICMP and TCP) to identify active hosts exist on the network, which can accurately locate the position of potential destination; and use TCP and UDP port to scan and detect the destination operating system and the type of service enabled. By scanning and snooping, an attacker can generally know the types of services provided by the system and potential security vulnerabilities, thus preparing for further intrusion into destination system

Malformed packet attacks

Malformed packet attack will send defective IP packet(such as fragment overlapping IP packets, TCP packets with illegal flags to the destination system, so that the destination system will crash and suffer damagewhen processing such IP packets. The main malformed packet attacks contain Ping of Death, Teardrop etc.

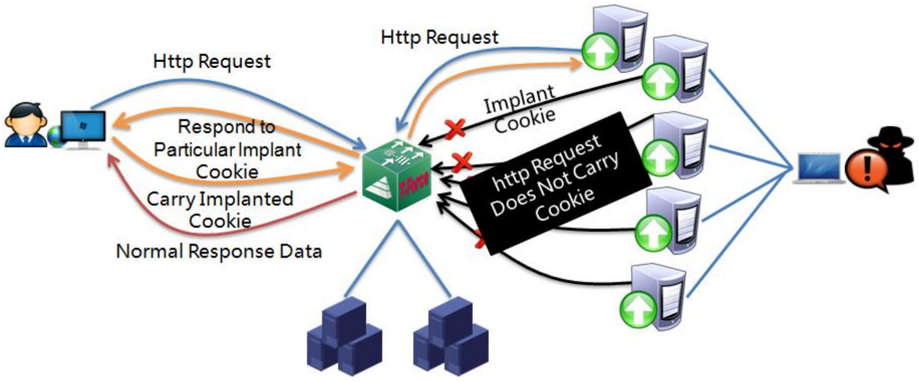
Venusense can effectively resist the common network attack packets via packet inspection, packet rate, number of connections limit function such as:

SynFlood/Jolt2/Land-base/ping of death/Tear drop/winnuke/smurf/TCP flag/ARP attacks/scanning TCP/UDP scanning/ping scan.

7.3.2 Application-layer DDoS protection

Application layer DoS attack is a kind ofattack mode that combines with the upper layer service. At present, the most common attack is the HTTP Flood attack (eg: CC attack). Compared with the traditional network-based layer DoS attacks, the consequence of application layer DoS attacks is quite significant and difficult to detect. HTTP Flood refers to attacksfrequently request resources from Web server by one or more clients, which will result in denial of service. Some Web pages are generally consuming much server resources, such as some resources are need to query the database or perform complex calculations. If such resources are requestedfrequently, it will make the server keep in busy state, thus realize the purpose of denial of service.

For the HTTP Flood attacks, Venusense can effectively identify attack behavior and normal request. If the Web server is suffered from HTTP Flood attacks, it will filter attacks, inhibit abnormal user from consuming Web server resource and will respond to normal requests, thus ensuring the availability and continuity of Web services. After users enable anti-CC attack, VenusenseADC will dynamically insert dedicatedCookie into the HTTP response returned by the server, When the normal browser accesses it, the Cookie will be carried back, while the "puppet machine" proxy server cannot recognize the Cookie, and will not carry it in the next request, so that the application delivery can distinguish normal traffic and attack traffic, and the attack traffic will be directly discarded.



7.3.3 Application layer access control

Standard firewall realizes intrusion prevention through identifying and controlling source, destination IP addresses, protocols and interface. In this operating mode, the application layer data cannot be accurately parsed, and the depth analysis against request cannot be made by combining WEB system, and the attack against WEB port can easily break into the firewall protection through legal ports.

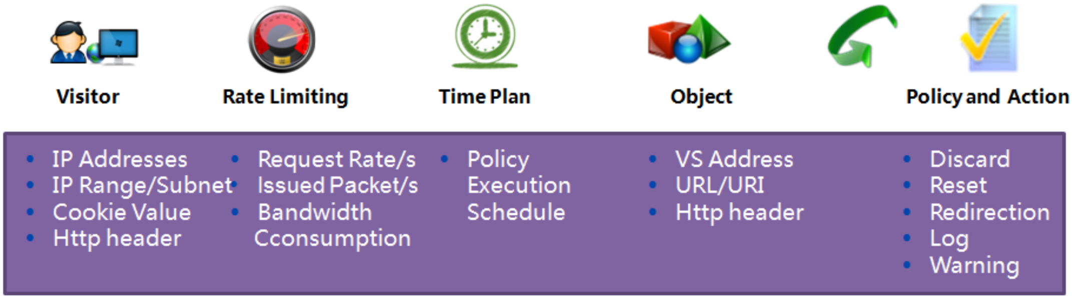
Access control based on just complement the shortage of network firewall. Compared with the traditional firewall, application-based access control is reflected in:

✧ Complete parse HTTP protocol, including HTTP headers, URL, Cookie, and provide HTTP protocol compliance check, avoid mixing illegal attack packets;

✧ Provides application-layer control rules to allow only legalusers to accessVenusense can achieveaccess control based on HTTP header, Cookie, request rate, VS address, URL. Prevent unauthorized access to the WEB.

✧ Implement restrictions on the URI, and only allows user to access the preset URI and prevent leakage of server resources;

✧ Implement traffic control on a specific URL, control Web page access frequency control according to the processing performance of the Web server, ensure that Web pagesthat consume relatively large resource can be accessed within the loading range of Web server's performance;



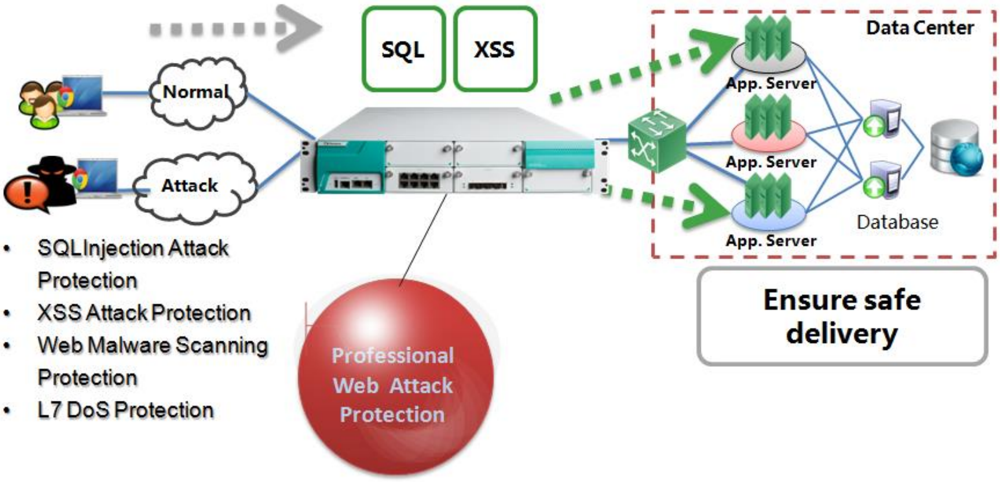
Visitor	Rate Limiting	Time Plan	Object	Policy and Action
• IP Addresses • IP Range/Subnet • Cookie Value • Http header	• Request Rate/s Issued Packet/s • Bandwidth Cconsumption	• Policy Execution Schedule	• VS Address • URL/URI • Http header	• Discard • Reset • Redirection • Log • Warning

7.3.4 SQL, XSS attack prevention

SQL injection attacks exploit Web application’s defects that do not filter and check the input data, and inject malicious SQL commands into the backend database engine to execute to steal data or even control the database server. XSS attacks refer to the malicious attackers insert malicious HTML code into the Web page. When the victims access the Web page, HTML code embedded into the page will be executed in the victim’s Web client, thus achieving malicious purposes.

Since what SQL injection and XXS attacks exploit are not common vulnerabilities but the individual defects of each Web page, the number of variants and modification attack is huge. In that case, the false negative and false positive rate will be extremely high if the usual modes are used to detect. With VXID patented technology, Venusense will analyze the modes of attack rather than code analysis feature, which can accurately and comprehensively detect and defend such Web attack.

VXID algorithm is divided into two stages: the first stage is the phase of behavior extraction, analyze and extract behavior features of Web attacks rather than data features, and establish Web attack behavior feature; the second stage is the real time analysis of network data. The "light virtual machine" established in the Venusense application delivery internally can simulate attacks behavior and observe their behavior features, and properly judge the occurrence of attack behavior. This detection mode based-on the principle can avoid high false negative rate caused by the matching with the curing features, and also avoid false alarms arising from the harsh detection rules.



VenusenseADC provides users with complete solution of application to realize safe and fast delivery service.. It not only minimizes the bandwidth but also ensures data security and forwards the traffic to the optimal performance of the application server via reasonable scheduling algorithm. This solution is different from the solution that combines with a plurality of product. The single deployment program can provide the best performance and optimized features on single devices and maximize the protection of user investment, thus providing fast, stable and reliable security experience for user’s key services.

7.3.5 Manage server connections

The maximum number of concurrent connections threshold for each application can be specified by VenusenseADC, which cannot only protect the application server resources, but also prevent certain applications from consuming too much system resources, resulting in normal service failing to provide toother applications.

7.3.6 Information leakage prevention

- HTTP header information hiding: VenusenseADC can erase specified information of response header in the server response packet, such as web server name, version number. It can also modify the real link directory of server to hide critical information of server. The allowed header can also be defined in the Response, whereas the other header information that does not satisfy the requirement will be deleted from the packet.
- Cookie Encryption: the plaintext Cookie information in the server response packets will be encrypted, and then sent to the client to prevent certain critical information(such as user accounts, online banking data) that is stored with Cookie from intercepting during transmission. In response to the client process, the Cookie information will be decrypted and sent to the application server, which facilitates unified management of application system and reduces the load of application system.

Chapter 8 Virtualization Technology

8.1 Overview

With the acceleration landing of cloud computing, large amounts of data and computing demand will be converged in the data center which creates unprecedented pressure over the data center. Moreover, the rapid development of mobile Internet provides not only rich applications to user, but also allows application system in the data center to become more complex and difficult to manage. In order to successfully use cloud platform, IT organizations shall have the ability to achieve agility of application migration. This process and the cloud platform itself is a new topic for most of enterprises.

As a bridge between the data center and application, Venusense ADC products provide effective integration with cloud platform to support virtualization linkage with Vmware, thus providing system hypervisor layer for data center server consolidation. In addition, Venusense ADC device’s virtualization feature makes up the gap in the corporate tenant cloud computing environments.

8.2 ADC Virtualization

In order to meet the needs of new service models, virtualization technology came into being. Virtualization refers to multiple logical virtual instances divided in the same physical device with the virtualization technology. The applications can run on independent space of each virtual instance without affecting each other, thus significantly improving operating efficiency of the physical device.

With the Venusense ADC products, it can deploy independent load policy for multiple service departments through virtualization deployment and also enables the enterprise network to adapt to the new service with its deployment flexibility.

8.2.1. Concept of virtual hosts

Concept of virtual hosts

What is a virtual host? That is the physical servers, operating systems and applications being "packaged" into onefile - the mobile virtual machine (VM).

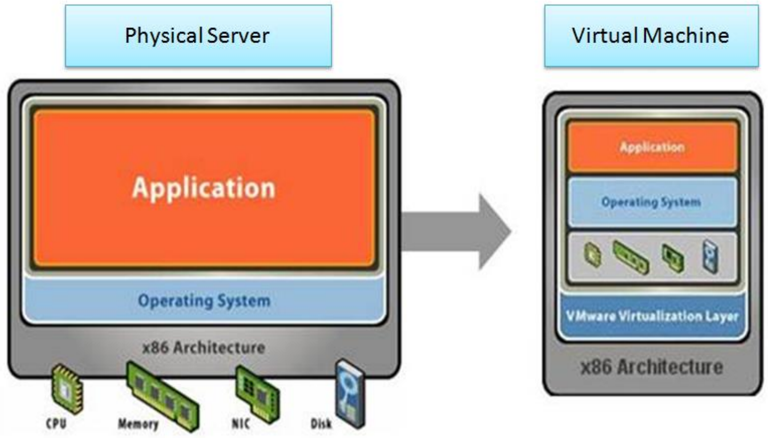


Figure 1computing virtualization - virtual host

In short, multiple virtual machines can run on a physical server, and underlying hardware will be shared by these virtual machines. From the application point of view, it is similar to a physical server that has its own operating system, cpu, memory, nic, storage and virtual resource. Users can independently manage individual virtual hosts, while the virtual hosts will not disturb each other, and run in their own service mode. Any virtual host can freely open, shut down, restart without affecting other virtual machines.

Venusense can simply load kernel module and convert system kernel module into a system hypervisor, and then derive a virtual device from the kernel module, which enables customers kernel mode (except traditional kernel mode and user mode). With virtual device, VM can separate address spaces from their kernel or any other running VM address space. After establishing VM, Venusense ADC’s guest operation systems can be activated in user space. Each guest operating system is a single process of host operating system (or system hypervisor).

Figure 2: The bottom is the hardware platform that has the ability to perform virtualization. The device running on the bare hardware is system admin host, which can achieve the creation and management of VM via the system admin during the process of achieving service in the VM.

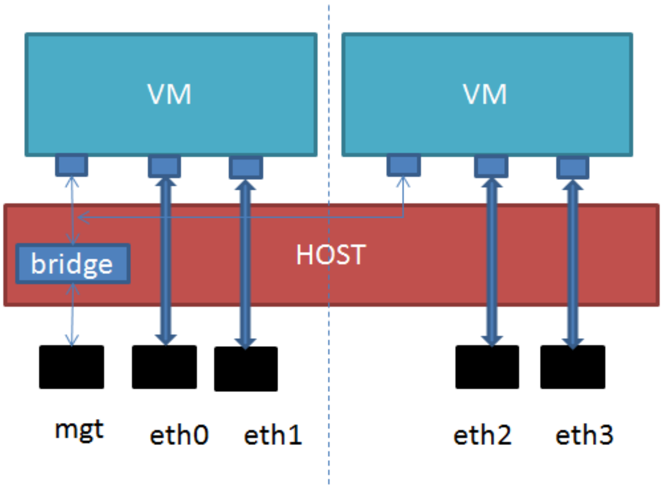


Figure 2VenusenseADC System Architecture

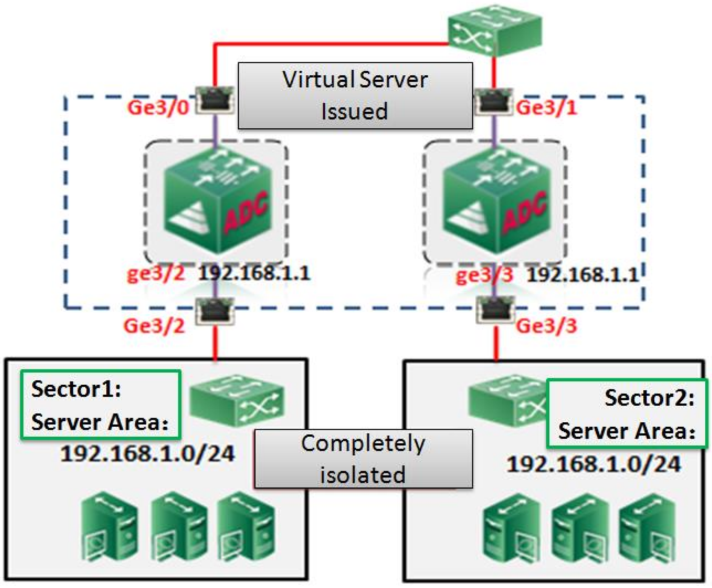
8.2.2. Venusense ADC virtualization features

Traditional shared mode virtualization will logically divide the physical host into multiple virtual hosts. Although the users can use the independent resources, CPU, memory, interface and other hardware resources are still shared between virtual hosts actually. When one of virtual hosts has high traffic or fails, it will cause the entire system unusable.

Unlike the virtualization technology of traditional sharing model, VenusenseADC supports hardware virtualization-based hypervisor:

- Up to 32 virtual hosts can be runningon a single physical server hosts at the same time;
- Each virtual machine on the same physical host is isolated;
- Each virtual host can specify its own CPU, memory and resource, and have an independent administrator;
- Each virtual host can run different service models, thus meeting customer needs to the greatest extent.

8.2.3. Venusense ADC virtualization benefits

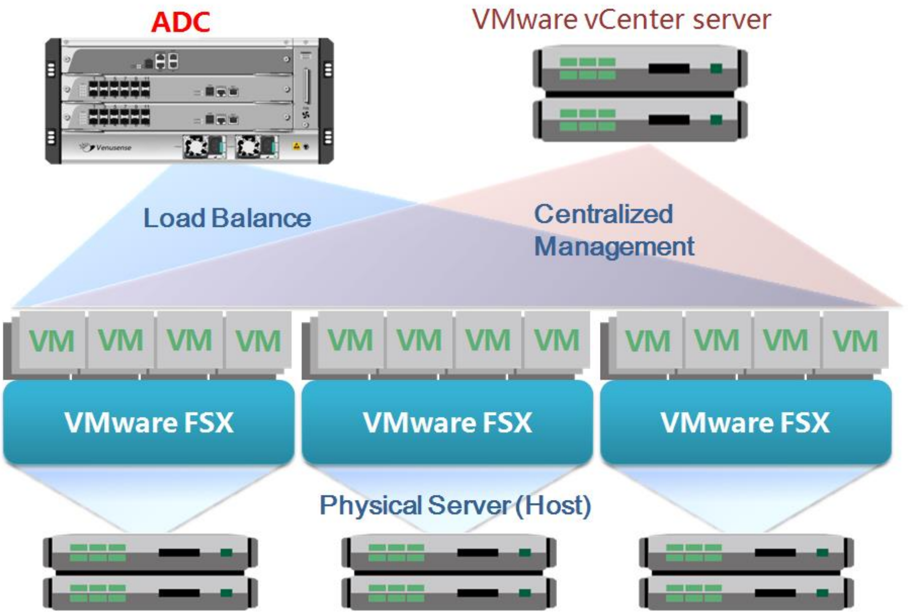


With the Venusense's virtualization technology, it can convert hardware resources into flexible service matching virtual service resources in a large-scale, thus flexible and effectively supporting future cloud data center applications.

- By deploying virtual host on a physical host, it can effectively solve the defects of traditional distribution deployments:
- Multiple enterprise users can reduce maintenance costs by simply deploying a separate physical host;
- The single physical host can greatly save space and energy costs;
- Network management is more simple, easy to maintain;

8.3. Linkage with Vmare virtual machine

Traditional data centers are moving toward the cloud platform that views the virtualization technology as the core. Venusense ADC closely follows service mainstream virtualization technology, and realizes linkage function with Vmware host. A solution with high availability and automation maintenance is presented for the application in the VMware vSphere architecture.

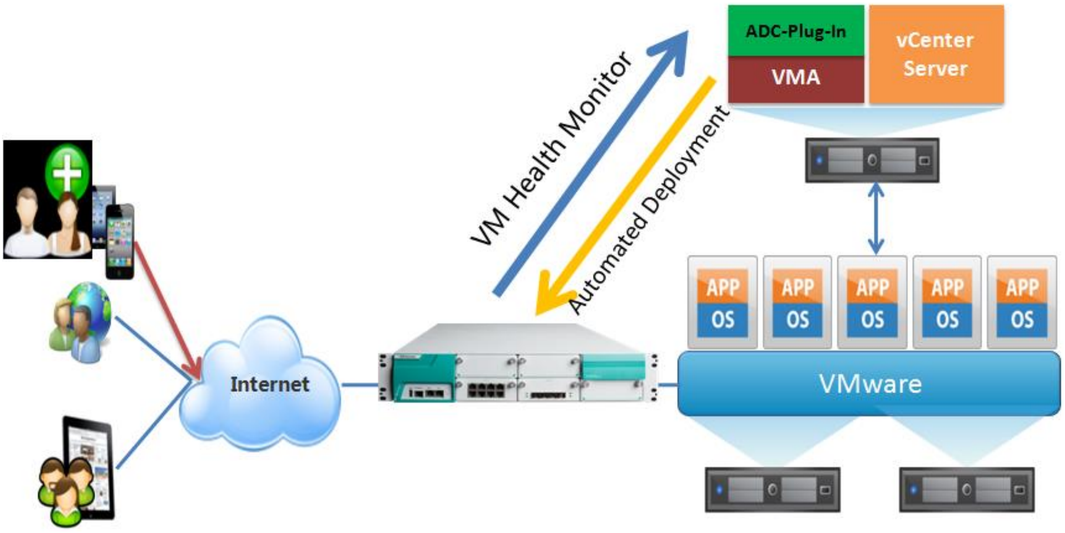


8.3.1. ADC technology and Vmware virtual host linkage technology

vCenter linkage is designed to provide solution for applications with high availability and automation under the VMware vSphere architecture.

By installing plug-in on VMA, plug-ins are communicating with vCenter to access the virtual machine information in the vSphere architectureon the one hand; and communicating with ADC to access server pool information on the ADC on the other hand.

By configuring the rules on the plug-ins and performing data analysis, the plug-ins candistribute the load balancing scheme to virtual machine to the ADC and realizes load balancing for service applications with ADC device; and also send instructions to the vCenter to let the vCenter uniformly manage virtual machine in the vSphere architecture.



8.3.2. Venusense ADC and Vmware linkage feature

By implementing the linkage with Vmware virtual host, it can automatically configure virtual machine on-line load balancing device. With the API interfaces provided by vSphere, it can communicate with the vCenter to access virtual machines deployment managed under vCenter inrealtime. If the new virtual machines (deploy, create and migrate via the template.) are created in the virtual environments and comply with discovery rules of pre-defined virtual machine, this virtual machine is automatically added to the server pool on this load balance to expand server cluster;

Perceive server cluster loadchanges, automatically manage offline and online virtual machine: automatically manage virtual machine’s suspension and switches via virtual machineCPU, memory, number of concurrent connections and health status of other service information and preset optimization rules. It can suspend some virtual machine, reducing the host load power consumption, energy conservation, reduce energy consumption and extend the life of the physical host if the service loadis relatively small; and will enable restricted virtual machines, automatically expand server cluster and protect the user experience if the traffic loadincreases

Perceive and repair the fault virtual machine to let it run again: determine the failure virtual machine and issue restart instruction to return to normal by analyzing the operation of the virtual machine, such as CPU, memory and service health status and other information.

8.3.3. Venusense ADC linkagewith Vmware

Compared with the traditional stand-alone deployment ADC products and linked with Vmware virtual machine host, Venusense ADC can automatically discover, deploy and remove the virtual machine host. It can significantly reduce server cluster expansion and the workload of operation and maintenance personnel. For traditional ADC products, if the virtual machine fails, the failed node shall be avoided scheduling. With Venusense ADC, it can perceive virtual machine, actively repair fault host, speed up the recovery rate of fault and effectively provide reliable service via analyzing the operation status of virtual machine.

Chapter 9 IPv6 technology

9.1. Overview

With the development of the Internet, the limitations of IPv4 become increasingly exposed, which severely restricts the application of IP technology and the development of future network; as the foundation for network of the next generation, IPv6 gains wide acceptance with its outstanding technology advantages. Predictably, in the future, the traditional IPv4 network will certainly be transferred to IPv6 and services of users will be faced with great challenges from IPv6.

Application delivery product of Venusense, through supporting IPv4 / IPv6 double protocol stack, may effectively connect the IPv4 and IPv6 network and decrease conflicts of user services in the migration of IPv6 network.

9.2. Defects of IPv4

The development and mixing of computer technology and communication technology makes the Internet application and scale develop rapidly. IPv4 technology gains huge success with its simplicity and effectiveness. However, the IPv4 agreement was drawn up in 1973, its early designers completely failed to anticipate that the IP network would reach the current development speed and scale. Defects and weakness crisis of IPv4 were gradually exposed in the 90s.

One of the biggest problems is the shortage of IP address resources. According to statistics, respectively 80%, 50% and 10% of class A, class B and class C addresses in IPv4 addresses were assigned in 1996. It is estimated by some experts that by 2010, IPv4 addresses may be in danger of running out. IP address is regarded as the ID of a network device node on the Internet. With the development of mobile and broadband technology, the IP address demand will be greater.

In addition to the problem of IP address, IPv4 is also faced with a series of problems such as large route table, Qos, mobile etc. In the route table, for example, IPv4 distributes addresses with the form irrelevant to the network topology, therefore, the number of route table increases rapidly with the increase of network. The huge route tables reduce both the efficiency of route nodes and the network and the stability of Internet services.

9.3. Advantage of IPv6

As early as 1990s, IPv6 was proposed for the improvement of IPv4. After 10 years of development, IPv6 is recognized as a future upgraded version of IPv4 technology.

As the next generation of network foundation, main advantages of IPv6 are:

Sufficient addresses - IPv6 is mainly aimed at the shortage of IPv4 addresses, of course, IPv6 has abundant resources of address. IPv6 absorbs the lessons of the IPv4 on insufficient address resources, the address length is expanded by 4 times, that is, 32-bit address of IPv4 is expanded to the 128-bit address of IPv6. This fully solves the problem of address depletion. If these IPv6 addresses were evenly distributed on the surface of the earth, each square meter of the earth may gain millions of IP addresses. Meanwhile, the scope of IPv6 address is limited. It still remains the concept similar to private addresses, which further increases the scalability of address application.

In the mean time of upgrading the IP, IPv6 also optimizes other parts of the IPv4 agreement:

Simple - simplifies the fixed basic header and improves the processing efficiency.

IPv6 simplifies the protocol header to improve the efficiency of network device for the processing of IP packet, such as cancels the IP header checksum field.

Strong scalability - the flexible extension header is introduced thus the agreement is easy to be extended.

IPv6 cancels the option, introduces a variety of extension headers, through which, in the mean time of improving the processing efficiency, it also greatly increases the flexibility of IPv6 and provides good expansibility to the IP protocol.

Level division - the address format is more hierarchical and is convenient for route aggregation.

The address space of IPv6 uses the hierarchical structure which is favorable for rapid search of the route. At the same time, with the help of route aggregation, it effectively reduces the size of the IPv6 route table.

Plug and play - the address configuration is simplified and the automatic configuration is achieved.

IPv6 introduces the automatic configuration and reconfiguration technology to automatically add， delete and update the configuration for information such as IP address to improve the manageability of IPv6.

Good safety - IPSec authentication and encryption of the network layer of end-to-end security

IPSec is originally designed for IPv6. IPv6 takes IPSec as the standard extension header of IPv6 and improves the end-to-end security feature.

Qos - new flow label field

To improve the congenital deficiency of IP, a flow label field is added in the header of IPv6 to improve characteristics of IP Qos.

Mobility - Mobile IPv6

By virtue of technology characteristics of IPv6, Mobile IPv6 better solves the IP mobility and is greatly improved compared with Mobile IPv4.

The above characteristics of IPv6 fully cater to the future development direction of IP network in integration and mixing. It also improves the operational manageability of IP network.

9.4. Deployment for upgrading from IPv4 to IPv6

Technological advantages of IPv6 are obvious, but the application of IPv6 is faced with an important problem, namely the deployment of IPv6 network. The current network is dominated by IPv4, upgrading all network device to IPv6 for one time will be a shock to all service. Therefore, a variety of transition technologies provided by Venusense ADC is a perfect solution for solving the problem of IPv6.

IPv6 deployment will roughly go through a gradual process. In the initial stage, several local scattered IPv6 islands will appear in the network ocean of IPv4. In order to keep communication, these islands connected with each other through the tunnel crossing IPv4; with the application of the IPv6 scale, the original islands are gradually polymerized into a backbone IPv6 Internet and form a situation of co-existing with IPv4. The IPv6 backbone allows introducing large amount of new services and gives full play of many advantages of IPv6. In order to achieve the mutual network resource access of IPv6 and IPv4, it also needs to transform the server to realize interconnection of v6 and v4; finally, the backbone network of IPv4 will be gradually shrunk to local islands. By tunnel connection, IPv6 occupies the dominant position and has the global connectivity.

IPv6 provides a lot of transition technologies to realize the above evolution process. These transition technologies are mainly focus on solving two types of problems:

IPv6 island communication technology: Realizing the communication between the IPv6 network and IPv4 network

IPv6 and IPv4 communication technology: Resource access of two networks

At present, the ADC product of Venusense solves the above problems mainly through the double stack mode.

Double stack: namely, upgrade the device to the IPv6 while retaining the IPv4 support and having the access to IPv6 and IPv4 device at the same time. Including double protocol stack support, the application program determines the type of protocol stack to be used according to the type of the address sent back to the DNS address resolution. ADC product of Venusense meets the demands of users to exchange different network protocol stacks and achieve seamless cohesion of two different protocol stacks to transfer from the IPv4 network environment to the IPv6 network environment gradually.

Across protocol conversion is classified into two types: NAT46 and NAT64:

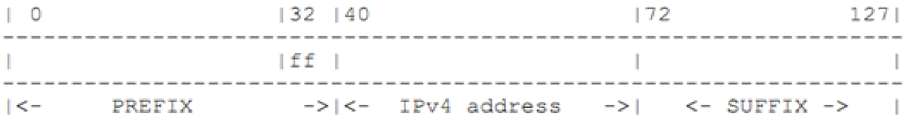
NAT46, namely, build a request from IPv4 terminal and transfer it into an IPv6 address;

NAT64, namely, build a request from IPv6 terminal and transfer it into an IPv4 address.

At present, Venusense ADC provides three transformation ways to achieve the visit between IPv4 network and IPv6 network: transition of IVI, embedded address and address pool. It is allowed to select proper transition mode according to the actual environment.

- **IVI conversion mode:**

IVI is a stateless address mapping proposed by China Education and Research Network (CERNET). Using the specified prefix can realize mutual transformation between IPv4 and IPv6 address. This mode supports NAT46 and NAT64.

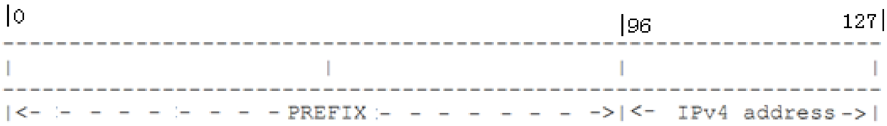


```
| 0                 |32 |40                 |72             127|
--------------------------------------------------------------------
|                   |ff |                   |                  |
--------------------------------------------------------------------
|<-      PREFIX       ->|<-  IPv4 address    ->|   <- SUFFIX ->   |
```

IVI address format

- **Mode of embedded address translation:**

The destination address after translation is a 32-bit address after the prefix of destination address of IPv6 is taken out according to the prefix of the user's configuration. The specific part of cutting depends on the prefix of the destination address of the user's configuration address (the prefix should not exceed 96 bits). This mode only supports NAT64.



```
|0
                                             |96            127|
--------------------------------------------------------------------
|                       |                    |                 |
--------------------------------------------------------------------
|<- :- - - - :- - - - PREFIX :- - - - - - - ->|<-  IPv4 address ->|
```

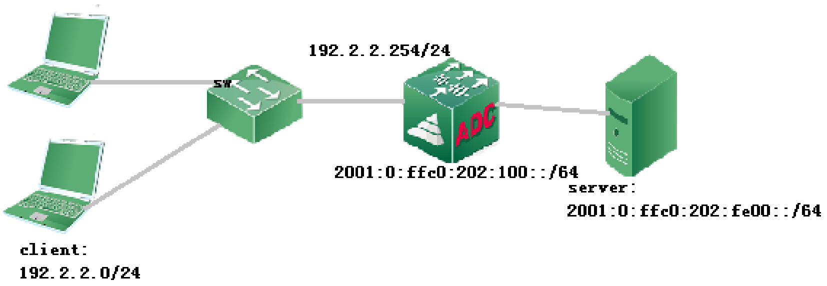
Embedded address format

- **Mode of address pool translation:**

This way refers to selecting the destination address after translation from the specified address pool. The original address may also be selected from the specified pool or be directly converted to the interface address. This mode supports NAT46 and NAT64 translation.

9.5. Typical deployment scenarios of ADC product

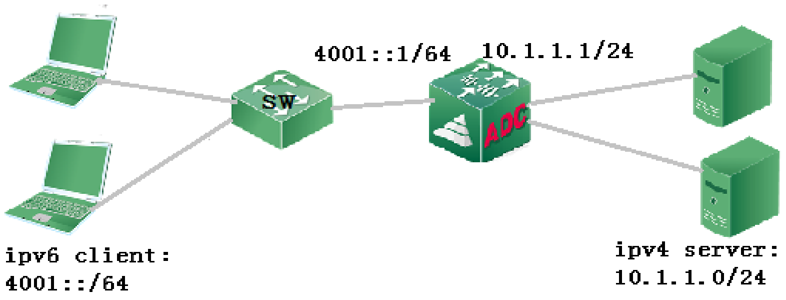
Typical conversion mode application scenarios of NAT46 IV:



Network environment:

IPv4 network on the public network needs to visit a server station in another type of IPv6 network through applications. Public IPV4 network segment is 192.2.2.0/24 and the address of the server site is 2001:0:ffco:202:fe00::/64. As the core router, the ADC is connected to the network in series.

Typical conversion mode application scenarios of NAT64 address pool:

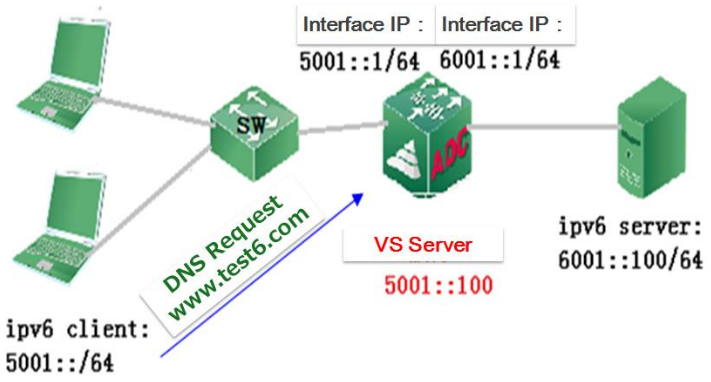


Network environment:

IPv6 network on the public network needs to visit a server cluster in another type of IPv4 network through applications.

Public IPv6 network segment is 4001::/64, the address network segment of the IPv4 server cluster is 10.1.1.0/24. As the core router, the ADC is connected to the network in series.

Typical application scenarios of NAT66+DNS6 :



Network environment:

Network users of IPv6 on the public network visit the server of another type of IPv6 network through domain name request.

IPv6 network segment of the public client 5001::/64 and the address of the IPv6 server is 6001::100/64. As the core router, the ADC is connected to the network in series.

The combination and application of different transitional technologies may provide a destinationIPv6 deployment scheme in combination with characteristics of the network itself.

For example, in general, the gateway device of a small home or office network might be upgraded to a double stack device to make use of the existing IPv4 connection with 6 to 4 access; as for enterprise network, edge devices might be upgraded to double stack devices. The pure link or tunnel mode might be applied to construct the IPv6 backbone for the internal core. The IPv6 connection with the branch directly uses link layer such as FR, ATM and the like to bear IPv6 services. For network ISPs, the existing access layer or convergence layer of IPv4 might be partially upgraded to double stack to support IPv6 network access. The existing MPLS backbone network might also be used for user access of IPv6. With the application and development of IPv6, the locally constructed pure IPv6 network may also be selected to provide special services such as IPv6 and the like and to be gradually mixed to the Internet of IPv6.

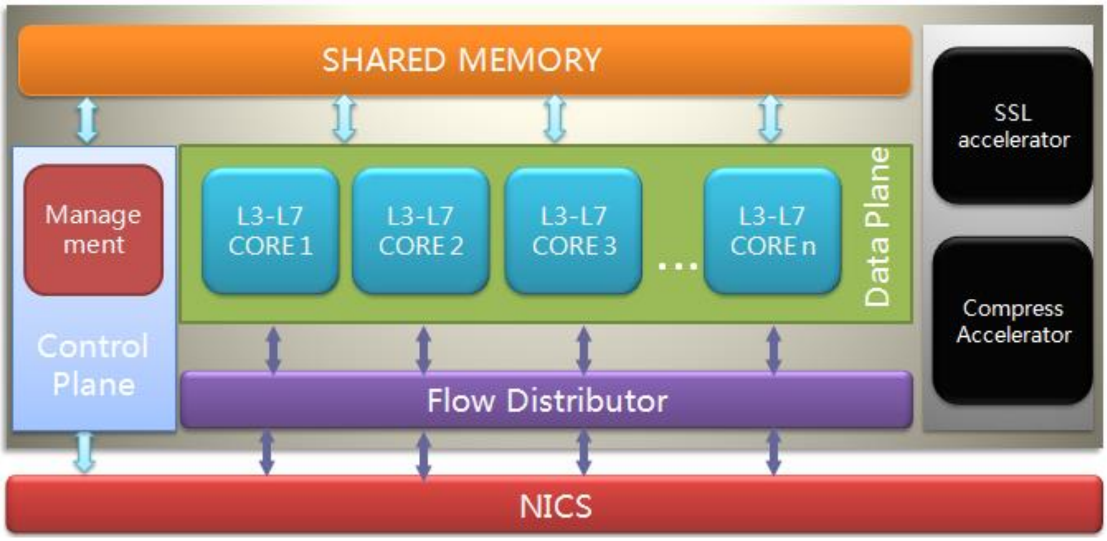
9.6. Development of IPv6

Looking from the future development of IPv6, the exhaustion of IPv4 address is still the main striving force for upgrading from IPv4 network to IPv6 network. At present, basic applications might be supplied by IPv6 are still confined to the range of IPv4. Although the prospect of IPv6 applications is very wide, these applications usually depend on the formation of a large scaled IPv6 network. However, the transitional deployment of the IPv6 network cannot be completed overnight. On a short view, there is a shortage of delighted applications and value added services of IPv6 to promote upgrading the IPv4 network to the IPv6 network. Of course, the commercial application is not the only driver for the development of IPv6. Other forces such as scientific research of the government are also promoting IPv6 forward continuously.

The full range the delivery series product of Venusense provides IPv6 and passes the IPv6 ready Phase-2 certification. Venusense has made full preparation to welcome the coming of IPv6 and has accumulated a long time for the IPv6 technology and related products of IPv6 to meet customer's demand of IPv6.

Chapter 10 Platform Advantage and Technical Innovation

10.1 T - Force platform architecture

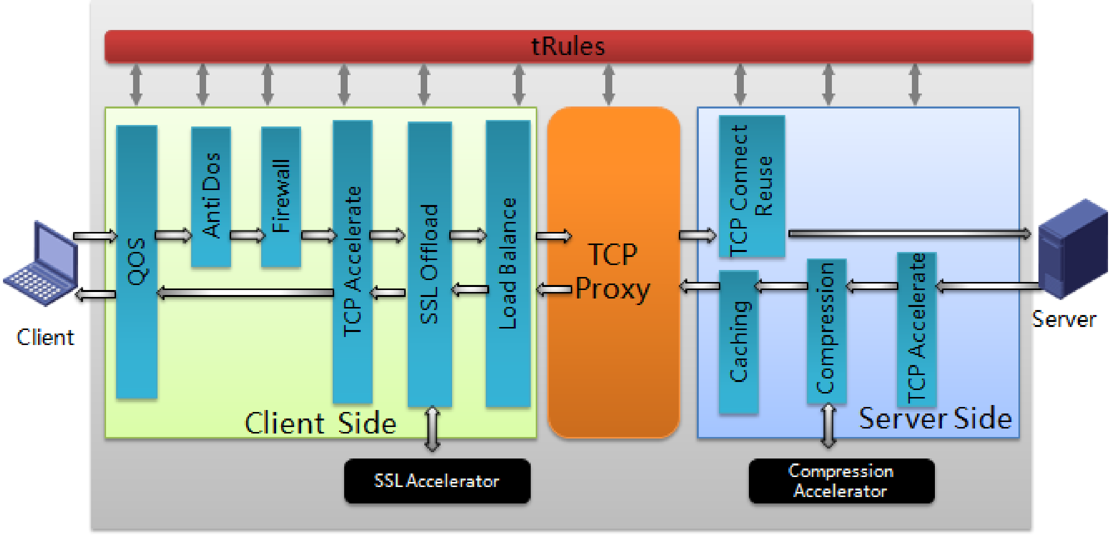


The multi-core CPU architecture is applied to all product of the T - Force ADC, which has the following core advantages:

- Full parallel multi-core technology: It can give the highest processing performance to play regardless of service types.
- **Hardware acceleration:** For SSL protocol offloadingand data compression which severely consume system resources, the Venusense platform uses specialized hardware to accelerate the engine,
- **Intelligent distributiontechnology:** The flow is evenly distributed to each CPU by a flow distributor. This ensures the balance of system resource usage.
- **Linear processing technology:** Through independent data processing of each CPU, it achieves high degree of linearization, minimizes the effects among CPUs and improves the system stability.
- **High performance engine for packet processing:** By optimizing the TCP/IP protocol stack, it achieves the zero copy and zero scheduling of message handling.
- **Highly scalability:** The processing performance of TBOS architecture presents a near-linear increase with the growth of CPU data.

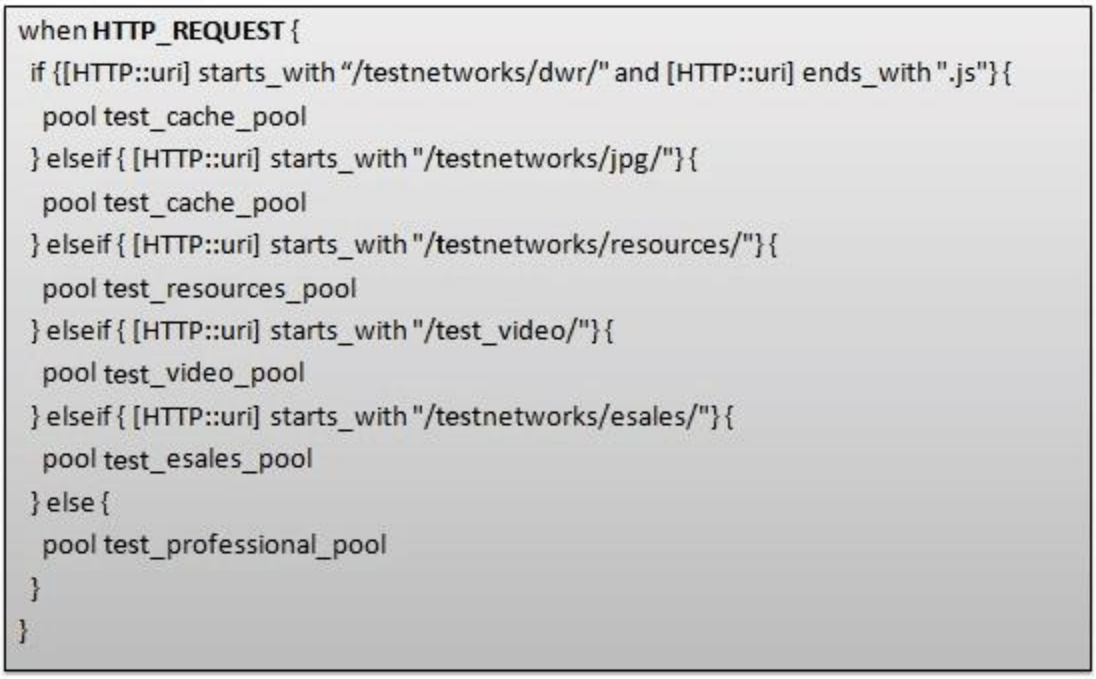
10.2 TBOS operating system

TBOS (T1 Balance Operate System) system is an operating system developed by Venusense Company for the VenusenseADC based on the profound technical accumulation. TBOS has fully independent intellectual property rights, integrates server load sharing, application acceleration, load sharing, multi-link load sharing, safety protection and other functions into an organic whole, and provides a function of unified configuration.



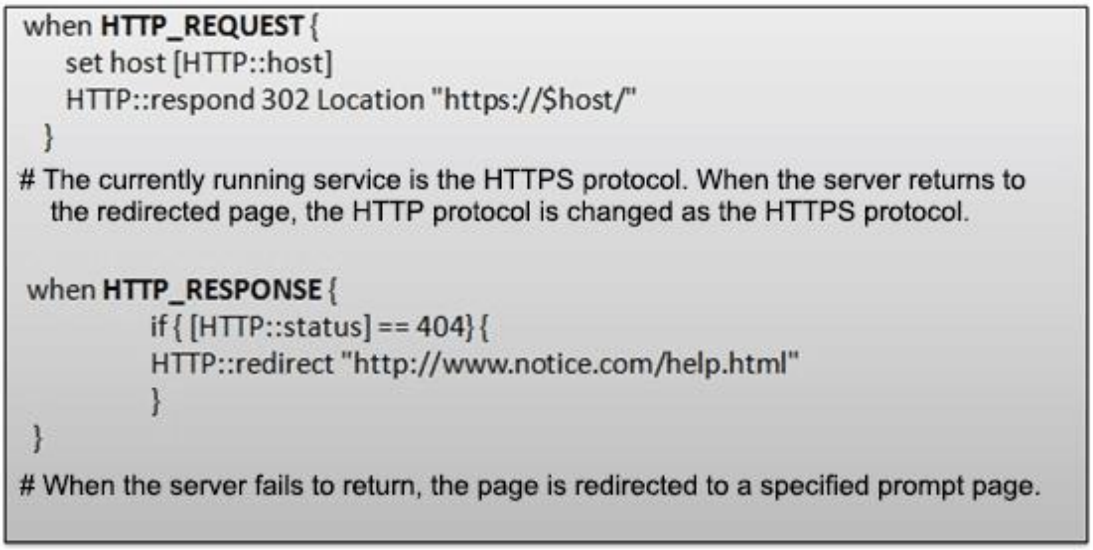
TRules is a powerful, programmable intelligent script strategy. Venusense made some development on standard TCL language. You may control each module and processing link of the TBOS system by the tRules programming to achieve precise control and personalized need of services to the maximum.

With the help of the tRules engine, you may decide the subsequent transferring after complex logic judgment of some data stream to achieve customized strategy which is unable to be completed on the graphical setting interface. For example: According to different HTML pages of client requests, the flow is distributed to the Cache server, resource server, image server, etc.



```
when HTTP_REQUEST {
 if {[HTTP::uri] starts_with “/testnetworks/dwr/" and [HTTP::uri] ends_with ".js"}{
  pool test_cache_pool
 } elseif { [HTTP::uri] starts_with "/testnetworks/jpg/"}{
  pool test_cache_pool
 } elseif { [HTTP::uri] starts_with "/testnetworks/resources/"}{
  pool test_resources_pool
 } elseif { [HTTP::uri] starts_with "/test_video/"}{
  pool test_video_pool
 } elseif { [HTTP::uri] starts_with "/testnetworks/esales/"}{
  pool test_esales_pool
 } else {
  pool test_professional_pool
 }
}
```

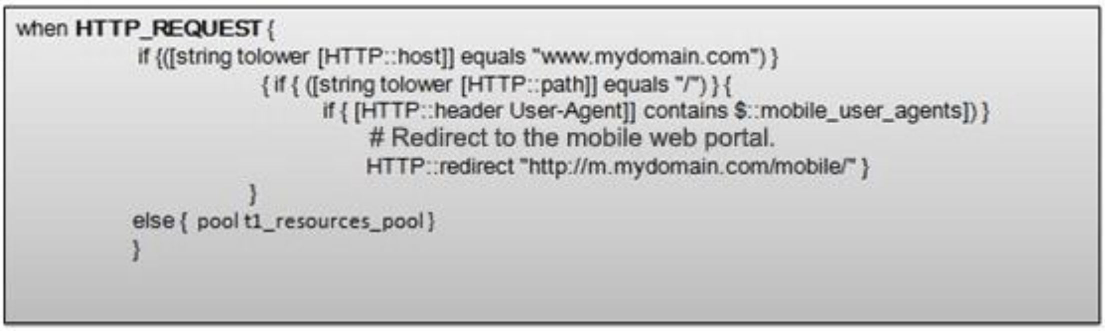
You can dynamically judge the abnormal page returned by the server and transfer to the preset page by the redirection technology to enhance the user experience:



```
when HTTP_REQUEST {
    set host [HTTP::host]
    HTTP::respond 302 Location "https://$host/"
  }
# The currently running service is the HTTPS protocol. When the server returns to
   the redirected page, the HTTP protocol is changed as the HTTPS protocol.

 when HTTP_RESPONSE {
            if { [HTTP::status] == 404} {
            HTTP::redirect "http://www.notice.com/help.html"
            }
 }
# When the server fails to return, the page is redirected to a specified prompt page.
```

You can judge that the client requesting access is a host or a smart-phone browser. In case of a smart-phone browser, the flow will be transferred to the servicer of mobile application.



```
when HTTP_REQUEST {
          if {([string tolower [HTTP::host]] equals "www.mydomain.com") }
                    { if { ([string tolower [HTTP::path]] equals "/") } {
                         if { [HTTP::header User-Agent]] contains $::mobile_user_agents]) }
                             # Redirect to the mobile web portal.
                             HTTP::redirect "http://m.mydomain.com/mobile/" }
                    }
          else {  pool t1_resources_pool }
          }
```