The network architecture optimization based on the server load balancing technology

Ruicai Huo*, *Songqiu* Liu, and *Shiwei* He

Beijing CSSC Information Technology Co. LTD, Beijing, 100094

Abstract. The paper takes the authentication gateway system for internal office network as the research object, for the performance bottleneck and single point of failure problem of the single authentication gateway deployed in the existing primary and standby modes, chooses an authentication gateway integration method based on CA certificate. The practical results show that: through implementing the authentication gateway cluster design and introducing load balancing mechanism, this method solves the performance, stability and single point of failure of the authentication gateway, and improves the resource utilization of the authentication gateway device. This method introduced in this paper can be used for reference for the network architecture optimization based on the server load balancing technology.

1 Introduction

In the early stage of the enterprise information construction, each application system is mostly deployed for single-node, and users can directly access the server address for one-to-one access. With the rapid development of enterprise scale and internet technology, users have more visits to the server application system, and the server response speed is also becoming faster. It's necessary to consider use multiple servers when a single server is unable to handle the many access due to insufficient performance, the way is load balancing.

Combined with the actual construction and operation of the application system in internal network. At present, the number of application systems and office staff have increased greatly, which will inevitably lead to the increase of visitors, the performance and stability of the application system are facing great risks. Because the background application server adopts the single node deployment mode, there's a single point of failure and great security risks. To ensure the safe and efficient operation, it's necessary to improve the deployment architecture of server and network equipment, adopts a load balancing technology to optimize the current network architecture, so as to better meet the users' requirements for current business.

2 Current situation description

At present, the internal network users and application system points belong to different

* Corresponding author: huoruicai1985@163.com

security area, the access control is through the firewall policy, identity authentication gateway concatenated in the middle of firewall and application system, provides the security services such as the unified identity authentication, access control, and the single sign-on for the entire network users in the way of double machine deployment. The identity authentication gateway connecting nearly 40 applications.The network deployment architecture is shown in the figure below:

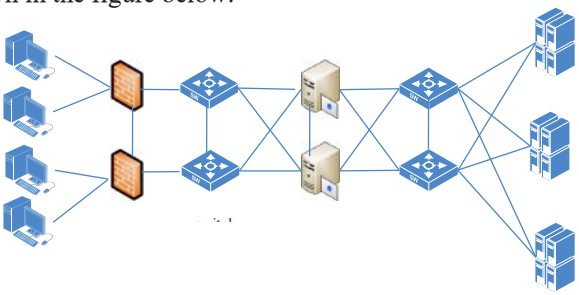

Fig. 1. The network architecture diagram.

The actual situation of current business applications is:

(1) the business expansion:the number of the network applications has increased from more than 10 in the initial period to nearly 40 at present(the number of applications refers to the number of applications actually configured in the gateway);

(2) performance bottleneck:the number of users increased from about 500 to more than 700, and the number of concurrent users increased to more than 500 at the peak. The large amount of visits and data resulted in the downtime of the application server.

(3) no application load:At present, there isn't relevant load balancing equipment in the network, so the access request of resource business cannot be properly scheduled.

(4) a single point of deployment: Each application system is currently deployed on a single machine, there's performance and single point of failure risk.

3 The existing problems

3.1 The challenges of System performance and stability

With the continuous development and expansion of the overall scale of the network and the pace of information construction is advancing continuously, it's expected in the next three to five years, the number of the network users will gradually increase from the current 700 to nearly 2000.The increase in the number of application systems and personnel leads to a significant increase in the number of concurrent connections and access bandwidth, which directly leads to the performance and stability of the application system.

3.2 The single hidden danger of Application server

All application servers adopt the single node deployment mode, this model has a single point of failure, and the probability will also increase synchronously with the increase of business access requests, the application system has a great security risk.

3.3 The security hazard of Gateway deployment

The gateway adopts the deployment mode of dual hot standby, as the number of business visits increases, once the service is interrupted due to a performance bottleneck, all

authentication requests will be synchronized to the standby gateway, which will also be interrupted due to excessive access pressure within a certain time limit. Therefore, it needs to consider a new extended deployment method to better meet the specific needs of the current business for identity authentication.

4 Network architecture optimization design

Introducing the load balancing technology to optimize the existing network architecture, focusing on the identity authentication gateway and backend application server equipment, load sharing in the identity authentication and application system, to improve performance and stability of application system, network and security equipment.

The optimal design of the network architecture is shown in the figure below (The ellipse annotation part is the architecture optimization part):

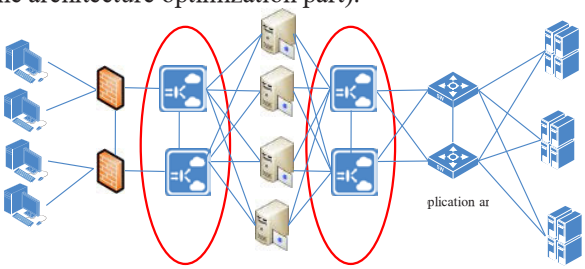

Fig. 2. The network architecture optimization diagram.

In single-arm mode, two load balancers are deployed respectively on the core switch before the authentication gateway cluster and the server cluster, without changing the existing network topology. The two groups of load balancer do high performance cluster.

4.1 Authentication gateway load

Under normal circumstances, two load balancers in the cluster can simultaneously load the user request of the back-end authentication gateway, which dynamically processes and allocates all data traffic on ports 80 and 443 based on how busy the back-end authentication server is and how many requests the load balancer itself is carrying, to realize the authentication gateway cluster and the load balancing cluster in the balanced utilization of equipment resources. When one device in the load balancing equipment cluster fails, the other device will be directly responsible for taking over all the traffic to the back-end authentication gateway, so that the business will not be interrupted. When there is a device failure in the back-end authentication gateway cluster, the load cluster will dispatch the traffic of all users to the normal authentication gateway, to achieve zero service interruption.

The load balancer cluster will configure all real authentication gateways as virtual services to realize load balancer and directly publish a virtual service IP. At the same time, the load balancer can continuously check the health status of the authentication gateway and remove it from the load balancer group once the authentication gateway is found.

4.2 Application load

Under normal circumstances, two load balancers in the cluster can simultaneously load the user requests of the back-end applications, which dynamically process and allocate data

traffic based on how busy the back-end application server is and how many requests the load balancer itself can handle, to realize the balanced utilization of application server resources. When one device in the load balancing equipment cluster fails, the other device will be directly responsible for taking over all the traffic to the back-end application server, so that the business will not be interrupted. When there is a device failure in the back-end application server cluster, the load cluster will dispatch the traffic of all users to the normal application server to achieve zero business interruption.

The load balancer cluster will configure all real application servers as virtual services to achieve load balancer and directly publish a virtual service IP. At the same time, the load balancer checks the health of the application servers sustainably and removes them from the load balancer group once a failed application server is found.

5 Application access flow

After the deployment of the load balancer, the whole process of the user accessing the application system through the authentication gateway and load balancer (taking the integrated management platform as an example) is described as follows:

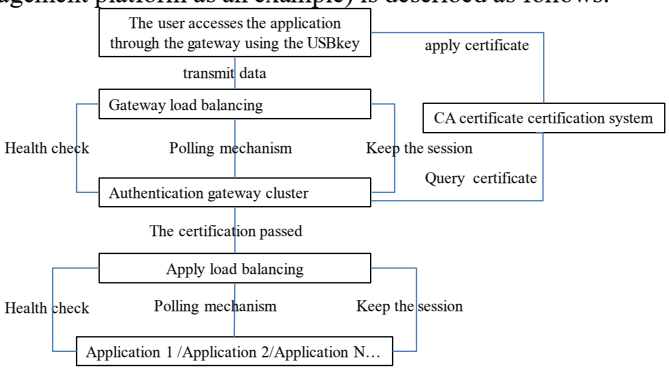

Fig. 3. System access flow chart.

(1) The user accesses the authentication gateway URL address through the browser;

(2) When the user request is sent to the load balancer, which reasonably allocates it to the corresponding authentication gateway according to the preset load policy;

(3) And the load balancer monitors the health of the authentication gateway in real time, which can find the fault gateway, and switch the user's access request to other normal authentication gateways in time;

(4) Through the identity authentication gateway load balancing strategy chosen, after receiving the user access request, requires the user to produce a digital certificate, the user input PIN code certificate in the web of box, identity authentication gateway read the certificate information, and user identity authentication, authenticated, pop-up gateway portal, click on links to integrated management platform system;

(5) When gateway certification through messages (traffic) to load balancing device, according to a predefined load balancers load strategy and reasonable distribution of the flow rate to one integrated management platform system, (if there are three sets of integrated management platform system, can be integrated management platform 1 2 \ \ integrated management platform integrated management platform 3 in any one);

(6) And the load balancing equipment can monitor the health of the integrated management platform system in real time, and can find the fault of the integrated management platform system, and timely switch the user's access request to other normal integrated management platform system;

(7) The integrated management platform system opens corresponding pages to users according to their rights, after receiving the certificate information and authentication passing information from the gateway.

6 Features of the optimized network architecture

Two-tier architecture deployment: the load balancer is deployed with a two-level architecture. The upper mainly for identity authentication gateway while the lower mainly for application system. Considering the single point of failure of the load balancer, two load balancer devices are deployed on the upper and lower main roads respectively (doing double active deployment) to ensure the high availability of the service.

Network architecture extensibility: the network architecture has good scalability, when an authentication gateway or application has a performance bottleneck, simply add physical devices to implement performance scaling without changing the original network architecture. The current load balancing scheme is only designed to provide load sharing services for authentication gateways and application systems. With the increase of application scenarios, load balancing devices can also provide load sharing services for more security devices, network devices and more application systems, so as to realize the network optimization of the whole network.

7 The network architecture optimization effect

(1) The network architecture was optimized by deploying load balancing equipment;

(2) Improve the resource utilization of authentication application system, network equipment and security equipment, to ensure the rapidity and stability of user access;

(3) The application or device in question can be switched to a normal through the detection mechanism, to avoid the single point of failure of the application system.

References

1. Gong yu Mo.Application of the load balancing technology in government office system[J] Digital technology & Application,2020
2. Shanshan Li,Cen Gao,Meiji Wang,Dongmei Li,Yanfei Jiao.Research on the application of LVS Resource load Strategy[J]. Computer system application.2019
3. Zhijun Shen.Research on high performance Switching technology and simulation based on load balancing structure[M]. Xidian University Press,2018
4. Kun Liu.Research on load balancing mechanism of cloud computing[M]. China Agricultural University Press,2018
5. Weiwe iLi, Zhaokuan Sun, Zhenyu Liang, Yan Hui. Research and application of server cluster technology[J]. Information Systems Engineering,2018
6. Zexing Shen,Yunjian Peng, Xishun Yue.Load balancing optimization algorithm for cluster servers under mixed requests[J]. Computer Engineering and Applications,2018
7. Yanrong Kong. Research and improvement of load balancing strategy based on Nginx high concurrency Web server[D]. Chang'an University,2018
8. Xin Wu. Application of load balancing mechanism based on virtualization in scheduling management business[J]. Power & Energy, 2018