



Load Balancing Strategies at Commerzbank

Traffic police on continuous duty



If too many people rush to the same door, it can become pretty tight squeezing out. If they have several doors to exit from, everyone can get out orderly, without waiting. This truism and ideal scenario should also apply to data traffic. It does at Commerzbank, where a concept for traffic management has been successfully implemented company wide.

The topic of load balancing first came to the attention of Commerzbank at the end of 1999. At that time, more and more locations were being equipped with Internet access. Initially, only the new high-rise building in downtown Frankfurt was online. "In the beginning, the access load was handled by an NT proxy server," explains Bernhard Heinz. His department is part of Commerzbank's Information Technology Production central service division and is responsible for all of the firm's network services. Their top priority is security, including everything related to firewalls and proxies. "When a single server no longer did the job, we set up a second one. Because the number of users – initially about a thousand – increased rapidly, we were soon dealing with a farm of 20 NT proxies, each with an identical configuration." Network specialists for Commerzbank had finally reached the point where they had no other option but to take a systematic approach to load distribution. The first step was simply to distribute outgoing Internet user requests (client load balancing). External access to the Commerzbank Web site was addressed later.

Two-pronged approach

The quest for the most appropriate concept proceeded in two directions. "At first, we considered a server-based load distribution system. We looked at the Windows NT Load Balancing Service (WLBS)," says Heinz. "The algorithm quickly reached its limits, so we decided against server-side load balancing." The drawback of this approach is that the load balancing area is very closely "wed" to the operating system and must be reset after every update. And even in newer versions it's not integrated as a core function but is delivered simply as an add-on. Simultaneously, the IT strategists followed a different path, which was to transfer load balancing to a separate device in order to eliminate dependence on the operating system. At the beginning of 2000, the network team became familiar with the BIG-IP products from F5 Networks. BIG-IP is a combined hardware/software appliance for data traffic management and application availability.

Test phase

A test scenario was configured to determine the behavior of the BIG-IP products in conjunction with two Internet proxy servers. As Heinz explains, "We didn't work with a load generator, but rather observed the load yield in what you could call a competitive environment." In up to 1000 Commerzbank branches in Germany – which corresponds to about 12,000 users – the test delivered solid information on load distribution, even under high traffic conditions.

In total, the Commerzbank's load distribution project "consumed" exactly six months. Six weeks were required to build the test environment, two months were needed for the IT experts to conduct internal tests, then came user tests in the production environment with consistently increasing load on the load balancer. During this phase no bottlenecks occurred on the side of the load balancing appliance. This continues to be the case. Commerzbank now has eight BIG-IP clusters (16 appliances) operating in Frankfurt, with three additional clusters in Singapore. The IT infrastructure in the bank's Asian-Pacific division was similarly consolidated, based on the example of the Frankfurt computer center. All in all, about 30,000 employees around the world have been "relieved" of their network services duties. As well as outgoing traffic, some of the bank's Intranet applications – such as SAP – have also been integrated into the load balancing concept.

Keeping good connections



One of BIG-IP's central tasks includes what is referred to as functional load balancing. "We process many sessions using the HTTPS protocol," explains Heinz. "Communication is dependent on the client communicating with the same server during a session." HTTPS connections function with SSL (Secure Socket Layer) encoding. However, establishing a session requires a substantial computing effort. Client and server exchange security certificates and determine encoding and compression methods. If a session is terminated, the entire procedure is executed anew. What this means for the load of the server CPU when several thousand users want simultaneous access is not difficult to imagine. SSL session persistence also has the task of maintaining a single connection between client and server within a defined time interval. The recognition attribute for the load balancer is the SSL session ID, which is assigned at the time of the connection and

uniquely identifies client and server as partners for the current connection. SSL session persistence is very important, for example, in many e-commerce applications in which client information necessary for a transaction is located only on the target server of the initial session but not dynamically forwarded to the other servers. If, while "shopping" around several Web sites, the user is always directed via valid load balancing processes to a different server where information from the original session is not available, the transactions are never completed. At this point, the persistence algorithms technically overwrite the general rules valid for load balancing. Furthermore, SSL persistence plays an additional important role for outgoing traffic. Many firewalls translate the addresses used by the internal network into one or more IP addresses, which are administered by the firewall. In this way, the firewall can pass traffic through to the Internet without divulging the IP addresses used internally. The return address in the request is that of the firewall. This is retranslated into the corresponding internal user address and forwarded to the protected network. If the port number is also translated, the firewall can use the same IP address for several users. This process is called "many-to-one NAT" or "address overloading" and allows the network behind the firewall to create thousands of connections to the Internet using a single IP address. Without persistence based on the SSL session ID, the correct assignment between client and original server could not occur.

Geographic load distribution

The 3-DNS Controller appeared on the scene at the beginning of 2003. This appliance administers and distributes user requests across several server sites. The requests are then forwarded, dependent on round-trip time, packet loss rates and other QoS (Quality of Service) criteria. The procedure is as follows: the client directs the request for a specific service to the local DNS (Domain Name System = system for naming computers and network services), which then forwards the request to the 3-DNS Controller. This acts as the primary DNS and clarifies the content and origin of the request. The controller then conducts a performance analysis of the BIG IP or individual server at each location. After evaluating the feedback from the various sites, the best-suited server for executing the client request is identified, i.e. 3-DNS responds with the appropriate IP address. If the service is unavailable anywhere, no address is issued. If services are available everywhere, the addresses are distributed based on the selected algorithm. "The health check based on the time-to-live value is particularly useful," says Heinz, describing Commerzbank's application of the 3-DNS system. "It identifies the number of hops – the connections between two network nodes – that a TCP/IP packet may retain until the target is reached. The host with the shortest connection then receives the request and generates a response." Another path used to conduct availability checks of services is the Layer 7 request. ICMP (Internet Control Message Protocol) determines whether or not a proxy is available in the network (the PING, for example, is based on ICMP echo request and response via ISO/OSI Layer 3). This does not yet mean, however, that the desired service is actually available. The application layer (Layer 7) test is the first to provide this information. The proxy is targeted by the F5 appliance with an HTTP-get request for a particular Internet site. If the request elicits a positive response, the service is available.

Making modification simple

Because an enterprise the size of Commerzbank is constantly confronted with expansion and restructuring of its various branches, the network management system must also be able to clearly follow the resulting modifications. "We can hide every internal service behind an IP address generated by BIG-IP," explains a Commerzbank network specialist, "and then switch it on and off, multiply it and distribute it to various locations within the bank. This requires no real effort." Another area has been more difficult, but here technological deficits are not the primary problem. "IT organisations in larger companies are characterized by the feature that various departments assume dedicated tasks," sums up Bernhard Heinz, "which makes a lot of sense. But if one team is responsible for load balancing and another for programming the Web server, there is always extra work involved in coordinating their activities when servers need to be temporarily or permanently removed from load distribution. Staff members first have to consult us. It would make much more sense if they could solve the problem on their own. But up until now there has always been an inability of the load balancing device in clientele processing." F5 now offers a solution to this problem – the iControl Services Manager. This allows all objects in the network that are administered with F5 devices, as well as appliances, to automatically switch on and off. The authorized user can modify the configuration for each object, examine relationships within the traffic management system using a central interface, and regroup objects. "If something like a Web server needs to be reprogrammed, the responsible staff member can remove the device from the network without involving us," explains Heinz.





F5 Networks – High availability from the traffic management specialist!

F5 Networks is an American manufacturer of traffic management solutions. The company's success is the result of its clearly directed technological specialization and the commitment of its employees. The Gartner Group considers F5 to be a market leader in the industry. F5 customers are enterprises that have stringent requirements for availability and system stability in their IT networks. Manufacturers like Microsoft, Oracle and BEA use F5 technologies to optimize their products. F5 develops and markets combined hardware and software solutions to address availability, security and performance of critical server applications in complex IT environments. Application traffic management is F5's expansion of the classic traffic management system, which has become necessary in the context of current database and ERP applications. F5 integrates Web services into this concept, making them more scalable and more secure.

"The services manager enters the modification in a system log so that we can always track the situation." This doesn't mean, however, that the network team can become remiss in their duties. In an expansive network, there is always a malfunction waiting to happen. However, a significant amount of progress has already been made when traffic can self-regulate on a large scale. Now, if only the ride home from work were so smooth...

F5 Networks Ltd

EMEA Headquarters
Clarke House, 65 High Street
Egham, Surrey TW20 9EY UK
Tel: +44 (0) 800 587 2233
Fax: + 44 (0) 1784 497 211

F5 Networks GmbH

Einsteinring 35
85609 München-Dornach, Germany
Tel: +49 (0) 89 94 383 222
Fax +49 (0) 89 94383 111

F5 Networks SARL

Le Sésame, 8 rue Germain Soufflot
78190 Montigny-le-Bretonneux
France
Tel: + 33 (0) 1 39 30 38 90
Fax: + 33 (0) 1 39 30 38 91

F5 Networks

Barcelona Business Center
Passeig de Gracia 24 -11
08007 Barcelona, Spain
Tel: + 34 (93) 228 78 00
Fax: + 34 (93) 228 78 99

F5 Networks

Hardwareweg 4, 3821 BM Amersfoort
Postbus 1466, 3800 BL Amersfoort
The Netherlands
Tel: +31 (0) 3345 466 11
Fax: +31 (0) 3345 466 66

F5 Networks

Liljekungsvägen 184
Hässelby S-165 75 Sweden
Tel: + 46 (0) 7057 820 21

www.f5.com