LANCOM SYSTEMS

White Paper

Load balancing

Modern enterprise applications have high bandwidth requirements. The ever increasing number of digital applications, such as HD streaming, cloud-service connectivity, and multimedia and voice applications, require bandwidths that, in many locations, are simply not available today. In other places, high-speed fiber-optic connections won't be ready for years to come. One solution is to use multiple Internet connections such as VDSL. The advantages: The load on the network can be intelligently distributed between the different connections and, in the event of a WAN link outage, access remains available via the secondary link. The technology that manages the efficient operation of multiple WAN links is called load balancing. This technology is used especially in modern SD-WAN infrastructures, for example to supplement or replace technically obsolete MPLS lines with multiple cost-effective Internet lines.

The concept

Load balancing distributes the data traffic evenly across the network. This ensures that data links are divided between at least two connections and, even if one link fails, the other one acts as a backup. TCP connections can be dynamically shared between the independent WAN connections by means of dynamic load balancing.

The network then has access to the sum of the bandwidths of the individual channels, although each individual TCP connection is limited to the bandwidth of the WAN connection assigned to it. Under some circumstances it may make sense to use static load balancing and assign bandwidth-intensive (e.g. HD streaming) or critical applications (e.g. Voice over IP) to dedicated WAN lines to ensure consistently high quality.

Packet-based vs. session-based load balancing

With packet-based load balancing, jitter may result in the transmitted packets arriving in a different order. Consequently, packet-based load balancing is unsuitable for many protocols.

This problem is avoided by session-based load balancing, which is recommended in these cases. For this reason, LANCOM only uses session-based load balancing.

Recommended devices

Typically, operating multiple Internet connections for load balancing requires the use of multiple devices. Routers usually have just one integrated modem, so additional modems have to be used. These scenarios therefore require several devices to be operated, maintained, and supplied with power. The increased number of devices results in a higher failure rate, too.

For this reason, the use of specially designed routers is recommended. These enable the simultaneous use of different types of Internet connection (DSL, fiber optic, 5G) by integrating multiple modems into just one device. This reduces the work involved in configuration because you only have to setup one device, and the probability of failure is reduced too.

For secure, multi-site data transfer, such as in branch networks, these devices should also support the IPSec-VPN encryption technology of the latest IKEv2 version.

The combination of these characteristics is referred to as a multi-WAN VPN gateway.

Backup vs. active/active (load balancing)

In the case of a backup, only one line is active and the second line is only activated if the primary connection fails. An example of this combines VDSL with an LTE/4G backup. Since 4G often goes hand-in-hand with a volume rate, this connection is not generally used for load balancing: Instead it should be available in the background to ensure high availability.

In contrast, load balancing in active/active operation mode relies on all of the connections being simultaneously active at all times. It makes full use of the existing bandwidth and the traffic can be distributed accordingly.

One side effect of load balancing is that you always have a backup.

Application scenarios

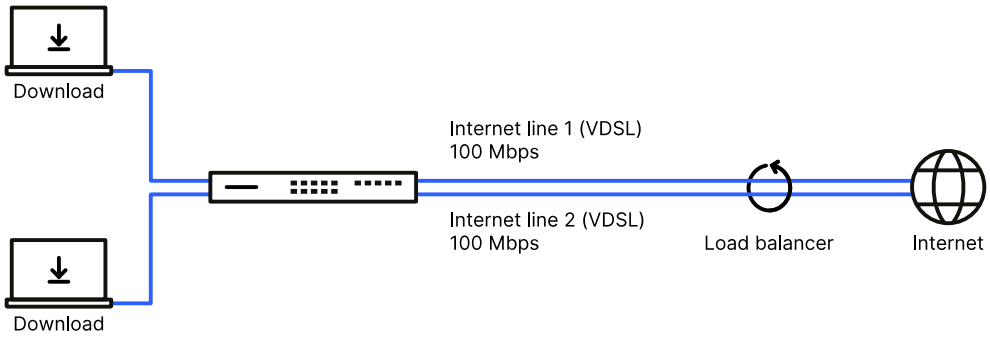

Figure 1:
Internet load balancing with two or more Clients

Two or more clients are connected to the Internet via a multi-WAN VPN gateway (2x 100 Mbps VDSL). With two clients connected, they are able to communicate with the Internet at the same speed (each at 100 Mbps) as the load balancer integrated into the gateway evenly distributes the data load between the two VDSL Internet lines. This ensures the maximum utilization of the available 200 Mbps for a significant increase in network efficiency.

This scenario works with all Internet connections. Furthermore, this feature only has to be supported by the local router, which means that there is no need for a separate gateway at the data center or main office.

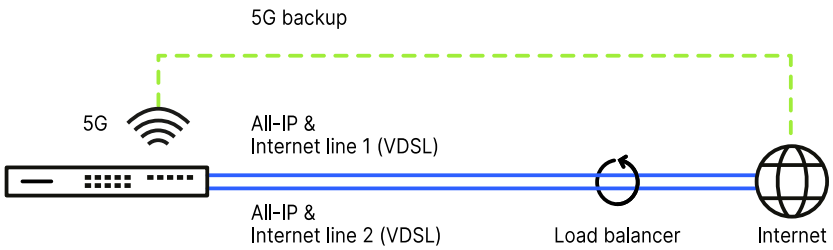

Figure 2:
Load balancing and VDSL/4G backup of Internet connections

Along with the efficiencies gained from load balancing, the failure of one Internet line is compensated by the other line, which can take over the data traffic. Special multi-WAN VPN gateways combine not only two VDSL modems but also an LTE/4G modem in a single device. This additional LTE/4G modem can be used as a backup connection in the event that the VDSL lines should fail—an ideal solution for maximum availability in mission-critical environments.

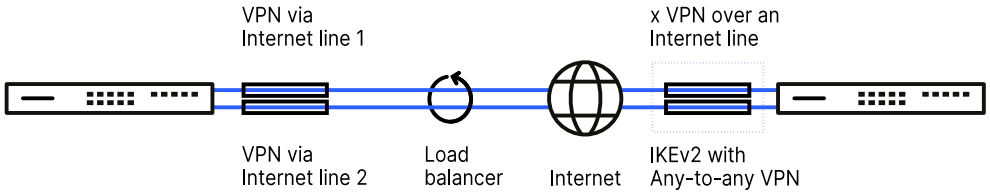

Figure 3:
VPN load balancing

IKEv2 VPN connections can also be combined to form a load balancer. The load of data traffic is more evenly balanced and the system has redundancy. The combination significantly increases the VPN bandwidth available for multiple clients.

In this scenario, both of the devices—the one at the branch and the one at the central site—must support the load balancing feature for VPN. It is important to ensure that the VPN connections in the load balancing network always terminate at the same target device at the central site.

Client binding

Due to the different NAT IP addresses of the servers, problems can arise when Internet load balancing is used for applications that use an IP address to identify the authenticated user (e.g. online banking). If a user is logged in to a web site, for example, and the load balancer then takes a different Internet connection, the server interprets this as a connection attempt by a user who has not yet been authenticated. In the best case the user sees a new login dialog, but not the desired web page.

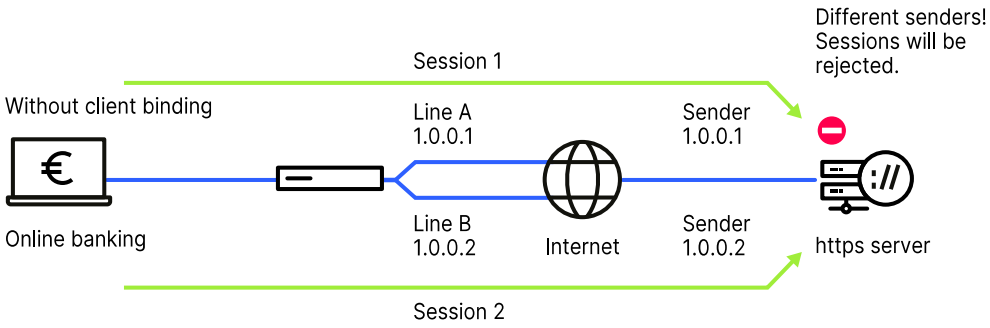

Figure 4:
Without client binding

One possible workaround would be to use a firewall rule (policy based routing) to direct the traffic to this server over a specific Internet connection. However, the full volume of the traffic from all the clients to that particular server would then be limited to the bandwidth of a single connection.

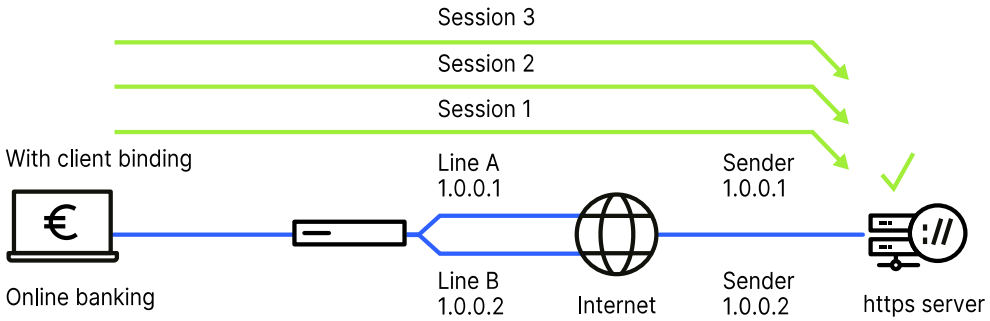

Figure 5:
With client binding

Also, configuring a backup becomes more complex.

In contrast to this, client binding monitors not the individual TCP/IP sessions but the client that opened the Internet connection in the initial session. All subsequent sessions from this client are directed over this Internet connection, a behavior that corresponds to the policy-based routing mentioned earlier. Client binding is protocol-related, i.e. it transports only data of the same protocol type (e.g. HTTPS) over this Internet connection. If the client loads additional data via an HTTP connection, it probably does this over a different connection.

Summary

This white paper served as an introduction to load balancing. This feature serves to increase bandwidth while spreading the load between the available connections. At the same time it introduces redundancy in addition to load balancing to provide high availability and, if desired, a backup connection can be made available as well.

LANCOM Systems GmbH
Adenauerstr. 20/B2
52146 Wuerselen | Germany
info@lancom.de
www.lancom-systems.com