WHITE PAPER

Everything You Need to Know About Network Failover

UP TO
16%
OF ANNUAL REVENUE IS LOST DUE TO NETWORK DOWNTIME

OVERVIEW

THE INTERNET has become so pervasive and integral for conducting business and communicating with customers, partners and employees, that network performance, high-availability, and uptime are now required for running the day-to-day operations of an organization. Network downtime not only costs money and loss of productivity, it can also adversely affect a company's reputation among customers and partners. For many companies, their entire business strategy depends on how well the network performs.

Many events can cause a network or site to go down, such as natural disasters, security attacks, a backhoe cutting a fiber optic cable, or failing network infrastructure. Organizations must plan for and allocate the budget to implement appropriate network infrastructure to ensure their datacenters and remote offices have the protection they need in anticipation of disasters.

According to market research firm Infonetics (now part of IHS, Inc.), large enterprises typically lose from one-half percent to 16 percent of their annual revenues due to network downtime. The more distributed a company's network is, the more likely it is to suffer service-provider interruptions. According to the survey, retailers are affected the most, with service providers accounting for more than 30 percent of their downtime costs. Another cause of downtime is human error, which accounts for about one-fifth of downtime costs. For financial institutions, this percentage jumps to nearly one-third.

This white paper describes the different types of failover that can mitigate downtime, the requirements of failover design, and strategies for successful failover implementation.

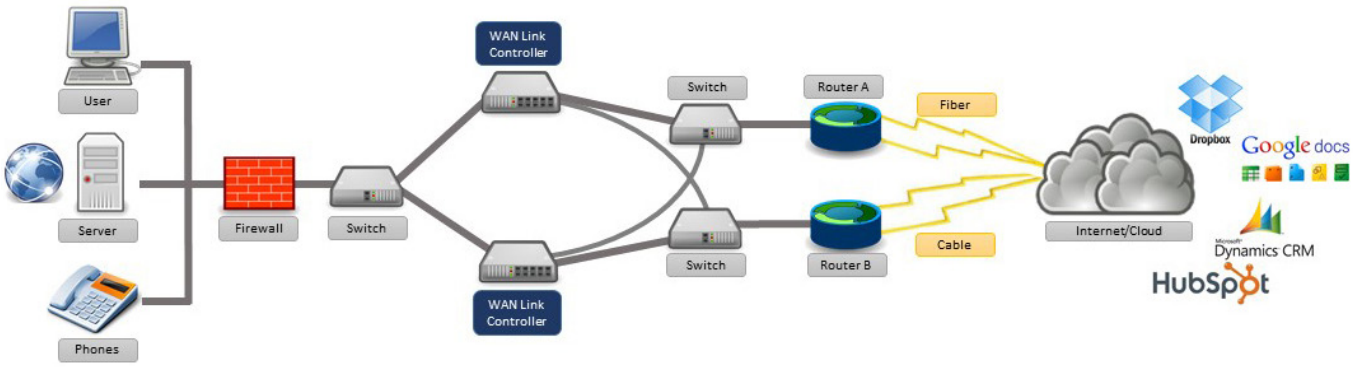
FAILOVER

Failover within a communications network is the process of instantly transferring tasks from a failed component to a similar redundant component to avoid disruption and maintain operations. Automated failover is the ability to quickly reroute data automatically from a failed component such as a server or network connection, to a functioning component, and is essential for mission-critical systems.

Different components may be configured for either cold standby (requiring human intervention), warm standby (automatic but delayed) or hot standby (automatic) failover. The three critical elements requiring failover configuration are power, network connectivity and server capacity.

DEVICE FAILOVER

In a failover situation involving a device, such as a firewall, router, WAN controller, server load balancer, disk drive, web server, etc., data is transferred to the same type of redundant component to ensure there is limited interruption in data flow and operation. If a primary component becomes unavailable because of either failure or scheduled downtime, the secondary component serves as a backup and takes over for its failed counterpart.



The capability to switch automatically to a redundant or standby system or network upon failure happens without human intervention (see Failover Heirarchy for other types of failover). Automated failover is essential in servers, systems or networks requiring continuous availability and a high degree of reliability — those that are responsible for mission-critical processes and data.

FAILOVER HIERARCHY

As mentioned earlier, there are different types of failover, some that are not entirely automatic by intention and require manual intervention. This is called “automated with manual approval”—activity is automatic once approval is given. When hardware is on “cold standby,” failover must be performed manually, which invites error.

In contrast, where hardware is on “warm standby,” the backup system runs in the background, so the transfer takes place automatically. The data on both systems is automatically synchronized. To the user, failover resembles a very fast automatic service reboot. However, the current transaction may be aborted

because it was not possible to synchronize the data prior to failure.

The most reliable failover scenario is "hot standby," where both systems permanently run in parallel — data on both systems is 100% synchronized at all times. Users will not be aware of any failures. This level of failover protection usually requires a corresponding modification to the client. To run both with systems in complete synchronicity, the connections to the client must be mirrored 100%. This normally requires clients that have connections with two or more servers at the same time, and can communicate with all of them. A normal web browser cannot do this.

Some enterprises implement hot failover and cold failover for disaster recovery. It is important to differentiate between failover and disaster recovery. Failover is a methodology to resume system availability in an acceptable period of time, while disaster recovery is a methodology to resume system availability when all failover strategies have failed.

CRITICAL ROLE OF FAILOVER

The convergence of voice, data, and video over a single IP network is making the network infrastructure one of the most critical elements in operational success. These Voice over IP (VoIP), video and data services are increasingly integrated with business-critical applications such as e-mail, customer relationship management (CRM), cloud-based storage, etc.

Therefore, all forms of communication with customers, suppliers and employees are inextricably tied to network operation. If the network fails, access to critical information can be lost or potentially compromised, with potentially calamitous results: for example, an airport risks massive delays that impact passengers, or patients' health may be compromised by a major medical center experiencing application delivery delays.

EXAMPLES OF ORGANIZATIONS THAT NEED FAILOVER

- Small and medium-sized businesses need inbound and outbound load balancing and failover services for an increasing assortment of critical-business traffic, from VoIP to email. For example, the local corner store that does online banking and bill-pay over the Internet, or a manufacturing company that needs email, web services, hosted ERP and ecommerce applications available all the time.

- Companies with a central headquarters and a number of branch offices and remote employees need secure and reliable data communications. They need reliable performance and high-availability of their VPN data, including the ability of the VPN connection to automatically failover if a WAN link goes down.

- Web hosting companies, MSPs, ASPs and small ISPs need incoming link aggregation and failover to ensure that their services are reliable, with extra bandwidth and redundancy available to their servers. Their mission-critical applications need to be up and running 24/7. If a WAN link goes down, the failover process has to be smooth and transparent to users.

- Many companies are deploying VoIP and Virtual Desktop Infrastructure (VDI) to cut expenses and enhance productivity. These companies now need quality of service levels and traffic-shaping for guaranteed bandwidth to critical services and applications such as VoIP and VDI.

- Companies that have ERP, CRM or any other software hosted in the cloud or accessed over the Internet.

FAILOVER REQUIREMENTS

Most corporate and government networks are comprised of three main elements — LAN, WAN and network infrastructure devices and services. The LAN provides interconnectivity around a single organizational location. The WAN provides interconnectivity between these locations (interconnecting specific geographical sites), other business partners, and access to public networks such as the public switched telephone network in the case of voice traffic, and the Internet for data traffic. The network infrastructure services element provides the services that allow control of the network and flow of data (DNS, DHCP, WINS, FTP), and contain access to the network using Active Directory, RADIUS, and TACACS, etc.

These three elements of network infrastructure services need the consideration of several requirements for creating a failover environment, the most basic of which is a connecting cable between the two devices. The second device initiates its systems only when it detects a problem in the first device. Some systems have the ability to page or send a message to a specific technician or support center. There may also be a third "spare parts" device that has running spare components for "hot" switching to prevent downtime.

The following are other critical elements that comprise a failover environment:

POWER

With power failures being one of the most common reasons for network and systems failures, all critical network components at either the primary datacenter, call center or failover site must be connected to a power source that has very high-availability — 99.999% in the case of a datacenter.

A LAN that provides critical services such as a hospital or bank should be equipped with uninterruptible power supplies (UPS) for each component of its distribution and access portions. These should be connected to emergency power sources to maintain internal communication. The WAN routers, switches, firewalls, etc. need the same form of protection to provide continuous communication and interconnection to external sites and other public networks.

Large datacenters and critical operations, such as call centers, must rely on multiple electric power companies to provide utility power to their locations. The power is brought into the critical site from different geographical locations. If power is interrupted by an accident that severs electric lines at a particular location, the other utility can continue to provide uninterrupted power.

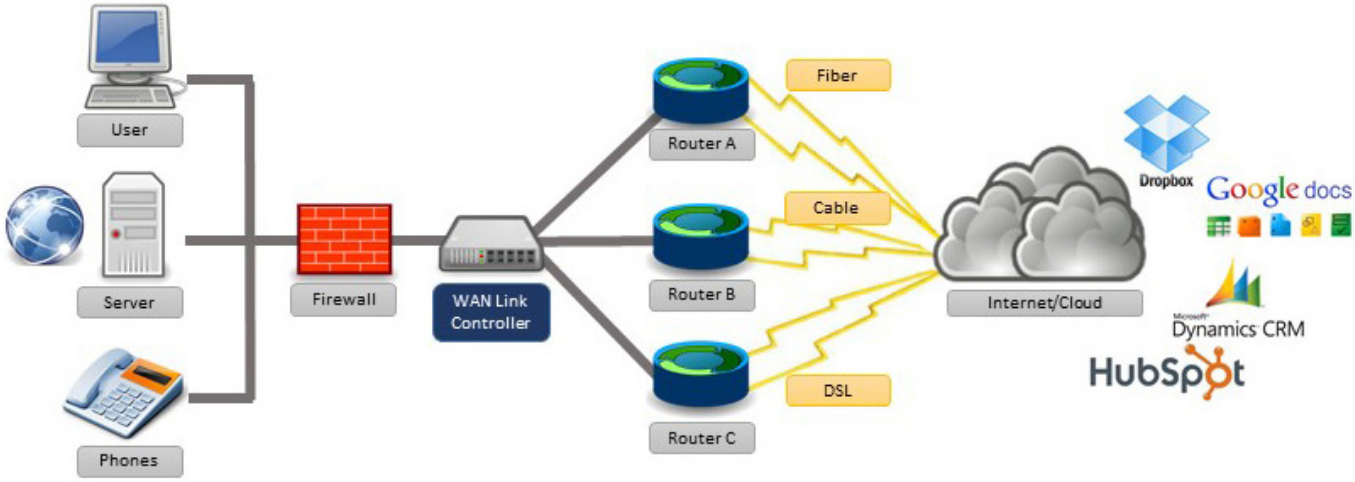
Emergency power generators may be used instead of alternate utilities. These generators, together with UPS equipment, can provide a continuous stream of electrical power for days if necessary, while utility power is being restored.

NETWORK REDUNDANCY

Levels of redundancy should be determined for the primary and backup networks based on the identification of critical network components, impact analyses and established recovery objectives. There should be consideration for redundancy of network devices such as switches, routers, gateways, etc. There should also be consideration given to redundant components such as power supplies, CPUs and circuit cards for the network switches and routers.

WAN LINK AGGREGATION

Consideration must be given to the redundancy and diversity of WAN links in conjunction with automated failover. Redundancy can be achieved by providing multiple links and multiple types of links for a single site and between multiple sites. For example, if the WAN network utilizes MPLS, it might be prudent to provide different links such as cable, so that if a carrier's entire service goes down, the organization can have a backup strategy, which may include satellite, cellular or microwave service.



Diversity of links can be accomplished either by link route diversity — two or more links travel different routes to your locations — or through carrier diversity. Multiple carriers are used to provide Internet access diversity and redundancy to companies that rely heavily on Internet connectivity.

WAN BANDWIDTH CAPACITY

Several capacity factors of alternate sites must be properly assessed in order to avoid failures caused by unanticipated high traffic volumes from a primary site. One factor is the peak capacity coming from the primary site that failed. The second factor is the peak capacity of the secondary site where the traffic will be rerouted to. The size of the WAN links should allow for both peak capacities, plus an additional 25-40% accommodating new peak traffic volumes. Additional traffic may come from new applications such as VoIP and other business applications, and/or traffic congestion caused by customers, suppliers, and employees.

Aggregated bandwidth should be ample enough to provide ISP failover and redundancy. If one link were to fail, you will still need enough bandwidth for users to be productive. Intelligent link load balancing monitors bandwidth availability throughout the network and priority-assigns traffic to the link with the greatest available bandwidth in order to guarantee that time-sensitive traffic (i.e. voice, video and other critical applications) receive the bandwidth required for smooth performance.

In addition to the availability of WAN links, there is a need for a mechanism to connect users to available servers. If a server where the user is connected suddenly becomes unavailable, the load balancer redirects the request to one of the other replicated servers. This action causes the loss of the original session-to-credential mapping where the user is new to this substitute server, and is normally forced to login again.

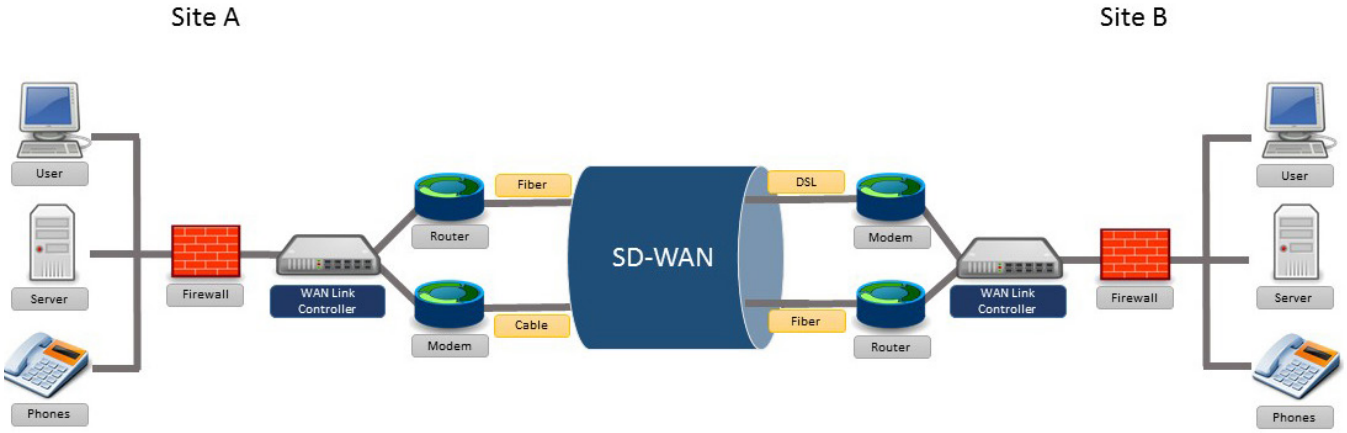
WAN LINK LOAD BALANCING AND FAILOVER

Many companies deploy WAN optimization solutions to merge WAN link load balancing and failover to cost-effectively eliminate downtime for business-critical, time-sensitive applications and ensure network performance. These devices enable redundant WAN and ISP access and can provide both outbound and inbound WAN/ISP load balancing and failover.

Bandwidth aggregation combines multiple WAN links into what is effectively one large network connection. Alternately, it can use bandwidth aggregation to maintain these links separately and allocate Internet traffic across them. Both techniques result in larger pools of available bandwidth and greater reliability.

SD-WAN AND WAN VIRTUALIZATION

For software-defined WAN features known as WAN Virtualization, devices with intelligent link load balancing are installed at both a local and remote site and direct traffic over the Internet between the two sites using the combined (or bonded) bandwidth of multiple ISP or WAN connections. Each site connected by such a bonded link is assigned a unique identifier that allows it to be differentiated from other sites.



Each site is also configured with addressing information for both the local and remote end of the bonded link. This allows the micro-appliance at each end to identify traffic that should be sent across the bonded link and direct it to the specified IP addresses on the WAN link(s) of the remote site. When the micro-appliance identifies such outgoing traffic, it is disassembled at the packet level into separate streams of data, then encapsulated for transmission through the virtualized WAN and sent over all available WAN links. Since each encapsulated packet contains addressing information for a specific remote location, data is easily reassembled at that location. For more information about SD-WAN functionality, please see Ecessa's technology reports at http://info.ecessa.com/what-is-sd-wan.

SUMMARY

The Internet has become essential for conducting business and communicating with customers, partners and employees. Network performance, reliability and uptime are essential for running the day-to-day operations of many organizations. Network downtime not only costs money and loss of productivity, it can also adversely affect a company's reputation among customers and partners. Many events can cause a network or site to go down, such as natural disasters, security attacks, human errors and infrastructure

failures. When evaluating how to avoid network failures, it is important to evaluate the many options available to ensure high-availability, network uptime and optimal network performance. It is also critical to examine solutions that will not only help avoid network failures, but are also affordable and will be operationally cost-effective.

KEY NETWORK FAILOVER OBJECTIVES

- Improve network performance and eliminate downtime for business-critical, time-sensitive applications
- Globally manage WAN resources
- Provide redundant hardware failover and monitoring capabilities for mission-critical applications to eliminate all potential single points of WAN link failure
- Establish reliable network connections
- Ensure inbound and outbound traffic management over best performing WAN link
- Provide WAN link load balancing for ample bandwidth for critical applications
- Cost-effectively increase scalability and throughput of WAN connectivity
- Failover to secondary datacenter if all links at primary datacenter are down
- For enterprises connecting many locations, coordinate SD-WAN among all locations to provide uninterrupted access to datacenter, Internet and cloud-hosted resources for reliable performance of applications such as VPN, VoIP, VDI, CRM, Office 365 and more.

ECESSA NETWORK FAILOVER SOLUTIONS

Ecessas meets key network failover objectives with affordable solutions for bandwidth aggregation, load balancing and failover that enable redundant WAN and ISP access and provide both outbound and inbound load balancing and failover to ensure uptime. Ecessa products optimize WAN infrastructure as defined by 24/7 availability, high-performance, flexible scalability and secure operations - while streamlining IT costs. They provide reliable access to datacenters and remote locations, ensuring business continuity when disaster strikes or WAN infrastructure is compromised.

ECESSA...

- Ensures each user gets the best network experience possible over the WAN
- Provides application high-availability over the WAN
- Directs traffic to only “available” WAN links and sites
- Enables administrators to optimize WAN traffic using cost-efficient WAN links
- Helps thwart network security threats
- Enables WAN link redundancy, ISP failover and Internet high-availability among multiple WAN links for important internal and customer applications
- Enables WAN Virtualization between multiple locations, providing uninterrupted communication for reliable performance of applications such as VPN, VoIP, VDI, etc.
- Provides redundant hardware failover and monitoring capabilities for mission-critical applications
- Provides traffic shaping and application prioritization capabilities for bandwidth management that guarantee your most critical applications get the bandwidth required for smooth and consistent performance
- Offers the best service and support in the industry, with 30 day free installation and configuration support and enhanced 24x7 warranty options.

13755 1st Avenue North, Plymouth, MN 55441 | www.ecessa.com
toll free: 1.800.669.6242 | 763.694.9949