In a 24x7 business environment,
the only good downtime
is no downtime.

THE IMPORTANCE OF NETWORK

The business network is the glue that holds almost every company together.

The network facilitates collaboration between departments, carries e-mail and phone calls to and from the outside world, connects sales people to their clients and to the people who will fulfill customers' orders, delivers training material and more. In fact, just about every business function in any decent-sized enterprise depends on the network.

That makes it safe to say that the network simply cannot go down — and if it does, it had better be back up quick. Most modern businesses aim for at least 99.9 percent uptime — that's three nines or roughly 9 hours of downtime a year.

That's actually well within the capability of today's networking hardware, provided the network is effectively architected and smartly managed. But in some industries, even that is too much downtime; hospitals, banks and retail establishments demand five nines or 99.999 percent uptime, which equates to just a few minutes of downtime a year.

RESILIENCE

The key to maximizing uptime is to combine resilient hardware with effective mechanisms that allow for effortless failing over to redundant systems and no interference with network service. This means carefully selecting and integrating network components, storage devices and servers that work well — and fail well — together.

Redundancy and Resiliency

Although redundancy is part of resilience, it's not the whole story. How redundancy is used and managed — as twell as the reliability of both the primary and redundant components — are also important considerations.

Rohit Mehra, director of enterprise communications infrastructure at research firm IDC, explains the difference like this: "You build redundancy and reliability into a product. Resilience is how you develop and build your network using those products."

The design aspect of resilience plays into a high-availability strategy. Although it might seem reasonable to simply install "two of everything" throughout the network, that's not advisable, Mehra says. "All redundancy has to be paid for."

This is not just in terms of financial cost but in terms of the management load on the IT staff. Simple duplication of the network, it turns out, can be a management challenge, creating monitoring and troubleshooting difficulty.

"With most networks, you're looking at N+1," Mehra notes. "That means for every number of switches or other components, you have one extra ready to go."

But the fact that a redundant component exists doesn't in itself guarantee resiliency — the reliability of that component is also a factor. Switching over to a power supply that is itself broken isn't resilience, nor is failing over to a redundant power supply if the component will have to be taken offline to repair the first, failed piece.

"If one piece fails, the other has to work long enough to replace it," says Dominic Wilde, director of marketing for HP Networking. "And the parts have to be hot-swappable, so you can replace a piece without bringing the device down."

Resilience emerges, then, from putting reliable components with "just enough" redundancy into a network configuration that is easy to manage and maintain. The more complex a network is, the more difficult it will be to manage and the more likely it becomes that human error could trigger an outage. >>>

The rise of server, storage and application virtualization is one way businesses have begun to reduce complexity. They tend to shift distributed business processes into the data center and under the watchful eye of IT managers using smart remote-management tools.

They can then prioritize vulnerabilities according to their potential for creating outages and alert IT personnel when to take action is another. Another crucial element: having policies in place that identify the different values of various applications on the network and the downtime threshold for each.

The average enterprise loses

3.6%

OF ITS ANNUAL REVENUE

to network downtime.

Moving Bits

At the root of network resilience is, of course, the network itself. This includes the switches, routers, cables, wireless interfaces, file and application servers, and everything else that transmits data from place to place. Designing for resilience encompasses three things:

- Eliminating single points of failure and/or network bottlenecks
- Decreasing complexity and the total number of components
- Establishing effective failover routes and connections

Total redundancy is not desirable because it leaves connections and components dormant, where failures can occur under the radar. Instead, at the connection level, an enterprise achieves resiliency by taking advantage of existing connections to route traffic around failures and chokepoints.

Spanning Tree Protocol (STP), Multiple Spanning Tree Protocol (MSTP) and Virtual Router Redundancy Protocol (VRRP) establish a standard for routers to find new pathways across the network and to redirect traffic around failed devices.

Because each protocol functions at specific layers of the network (VRRP for the network core, STP/MSTP for transport and endpoint access), the IT network team must implement and administer them in different ways. This can be difficult and time-consuming.

They also were not designed for use with increasingly popular communications services such as Voice over IP (VoIP) telephony and video conferencing. These demand low latency and low jitter. VoIP and video conferencing can trigger problems because of the lengthy reconvergence times (up to several seconds) and unoptimized traffic patterns they typically produce.

HP created the Intelligent Resilient Framework (IRF) in an attempt to address the shortcomings of older protocols through virtual switching. "IRF removes the need to have any of these protocols," says HP's Wilde. "You take two or more switches and effectively virtualize them so they're treated as a single switch."

IRF is an example of stacking: combining several switches into a single stack that a network administrator can then manage as a unique unit. Cisco Systems offers a similar capability, StackWise+. So far, there is no standard technology for switch stacking.

As Amy Wong, Cisco's marketing manager for borderless networks, says, "Different vendors' implementations exist and the level of stacking resiliency and manageability can vary widely from vendor to vendor. It's key to comb through some of the stacking elements — such as failover time, feature transparency in a stack, ease of implementation, scalability and manageability — to ensure that there's maximum resiliency and failover transparency in the network."

Both StackWise+ and IRF allow several switches (up to eight for IRF, nine for StackWise+) to be stacked and managed as a single unit. Because the redundant switches do not lay dormant until the primary switch fails but are actively involved in routing traffic through the network, failover happens almost instantaneously (under 50 milliseconds).

It's completely invisible to the rest of the network. As far as other devices are concerned, it's the same switch doing the same thing.

Sharing the Burden

Stacking can be an effective load-balancing tool, allowing each subcomponent to pick up the pace to distribute the burden of high usage across several switches. Implemented properly, load balancing solutions can provide almost instantaneous failover because the failure of a component merely shifts its workload to the rest of the already-active components.

Load balancing is simple in theory, if not always in practice. At the network level, it assigns traffic to the components best able to process and transfer it, assuring that traffic won't overwhelm any single component at any given time.

At the application level, load balancing consists of throttling bandwidth-greedy applications and optimizing their traffic. That way they can coexist with other network traffic.

"On every product and platform, load balancing has a different connotation," IDC's Mehra says. "In data centers, a load balancer is a separate device by itself. When you look at enterprise traffic patterns, load balancing is more around traffic routing.

The Number of "Nines" Needed

Every business has mission-critical applications they cannot afford to be without — even for just a few minutes. The key to smart high availability hinges on determining which apps require an outlay to assure five (or more) nines of uptime.

The cost of resilience rises exponentially. It can cost tens or hundreds or even thousands of times to gain another "nine" of uptime than what it cost to gain the last one. Moving from one nine to two (90 percent uptime to 99 percent) can be far, far less expensive than moving from four nines to five (99.99 percent to 99.999 percent).

Unfortunately, there's no hard and fast rule about how much uptime an organization will need. Different industries and companies have different needs, and the cost of outages can vary widely from one to the next.

"Markets such as healthcare, retail and manufacturing are all environments where your business requires the support of a network infrastructure," says Rohit Mehra, director of enterprise communications infrastructure for IDC. "You can't be a retail environment with a thousand stores and have your network go down."

The IT team and the business leaders must determine the financial value of an application to the business. For example, if a website generates $1 million dollars in sales a day and the database that records those sales is knocked offline for an hour, a company could theoretically lose 1/24 of that day's sales or more (depending on the time of day and whether the failure was during a peak sales hour), or about $40,000 (plus the cost of repair and IT staff-hours committed to restoring the connection).

In some cases, the uptime needed is dictated by regulatory structures. Sonal Shaw, associate partner at Solstice Consulting, points out that regulation severely limits banks. "You have a lot of time-sensitive transactions. If there's network latency and transactions cannot get back on time, there are legal consequences — including fines."

As a general rule though, the larger your enterprise, the more resilience you need. "It's more challenging in large businesses to do any kind of predictive patterning in terms of load structuring, environmental conditions, traffic patterns or mix of users," Mehra says. "So you have to build more resilience into the infrastructure."

"On a different level, you might have a wireless network in your enterprise, and you might be looking at load balancing based on the usage patterns and the number of users that might be on the network at any time," he adds. "You can use load balancing across almost all network components to build network resiliency."

Learn how CDW's network optimization solutions can improve network resilience and accessibility.

Data Availability

A redundant, reliable, always-up data connection without data to move over is not much use. This is why high availability storage proves crucial to a resilient network.

Numerous technologies exist to make redundant copies of a company's data available in case of hard drive failure. This includes redundant array of independent disks (RAID) striping within the disk array to simultaneous replication of every disk write to multiple locations both on — and offsite.

However, not all technologies for preventing data loss include adequate failover mechanisms that contribute to overall network resilience. A RAID-striped disk array will respond automatically to disk damage by reading the data from elsewhere in the array.

But if the array itself goes down or gets cut off from the network, RAID isn't much help. Offsite replication would be more useful, provided latency is not an issue, because local systems can be switched over to use remote servers fairly quickly.

For the most part though, storage area networks (SAN) provide the best option for high availability storage. A SAN pools available storage from across the network, apportioning it out to servers, applications and other devices as needed.

This maximizes the use of each physical drive, which can be provisioned to as many machines as needed. And, because the storage space allotted to any particular device is virtualized, data can be seamlessly moved between multiple physical drives with no interruption to the end user.

"With a SAN, you can add storage to a server or other device remotely, which is a godsend for administrators," says Laura DiDio, principal analyst at the tech research and consulting firm ITIC. "There's no downtime involved, and you often don't even have to take the server out of production."

Mission-Critical Network

With the trend toward greater virtualization putting emphasis on the network for even basic workplace tasks, network administrators are challenged more than ever to maintain ideal uptime. "Mission-critical networks are becoming more of the norm," says HP's Wilde.

Keeping those networks up and running is a challenge, but not an insurmountable one. Clever design and effective tools can create highly resilient networks. That is if IT chiefs forestall the urge to simply add more to address every possible problem.

According to Wilde, "The key to more availability is massive simplification. This includes the tools and infrastructure that make up the underling network." ◊