

To print: Select **File** and then **Print** from your browser's menu

This story was printed from ZDNet Asia .

10 ways to improve network performance

By Stephen Withers, Technology & Business magazine , Special to ZDNet Asia
25/2/2005

URL: <http://www.zdnetasia.com/insight/network/0,39044847,39218725,00.htm>



When your car starts to get sluggish you pop the bonnet and check the individual components underneath.

So why is it that when our networks start to run slow the components are often the last thing considered? We take the blame off the network itself by giving you 10 ways you can improve network performance.

When most people's idea of a network was a workgroup LAN, it either worked or it didn't. If it didn't, the most likely causes were a crashed server or a disconnected co-axial cable. Today's enterprise networks are far more complex, with more potential sources of failure or degradation, and a wide variety of different traffic types competing for scarce resources. One effect of this complexity is that there are many different components that can be managed, supplemented, replaced, or just plain tweaked in order to get the performance levels required. We trawled the industry and have come up with 10 ways you can improve network performance without having to replace or upgrade your infrastructure.

1. Understand your network

Without having an understanding of what's actually happening on your network, you are likely to fail at any attempt to address performance issues. Peter Prichard, marketing director Asia-Pacific at Compuware, says people tend to blame the network for poor performance, but the PCs and servers can also be the cause.

"The first thing to do is make sure the network really is the problem," Prichard says. "Even if it's not the network, IT spends a lot of time proving it's not." Tools such as Compuware's Vantage suite can isolate problems such as a slow client, excessive latency on a WAN link, or poorly written SQL on a back end server. An application might be developed on a LAN and then deployed over a WAN with disappointing results due to an excessive number of database calls. This sort of analysis may reveal things you didn't know about your network, such as a 1.5Mbps WAN link when you're paying for 2Mbps, says Peter Owen, territory manager at Packeteer. Collecting the right information also lets you take an active stance, identifying and dealing with problems before they impact on users.

Many people will blindly add bandwidth in an attempt to solve a perceived problem -- this tends to be one of the biggest mistakes people make, Prichard says. "You've got to have facts -- application-based facts," he says.

David Gibb, technical consultant with Vanco Australasia agrees. He says that what may dramatically improve performance in one environment could hinder performance in another.

Scott Atkinson, managed LAN services practice leader at Netforce, points out, there are a variety of free, cheap, and expensive tools that singly or in combination can show what's happening and why. MRTG (Multi Router Traffic Grapher), a free utility from <http://people.ee.ethz.ch/~oetiker/webtools/mrtg/>, is one that can help you gain an understanding of your network.

A network analyser itself will only show the aggregate traffic, and won't deliver the information you need. Prichard says to "start with the premise that the application is king", rather than checking individual aspects of the infrastructure.

Lorenzo Modesto, general manager at Bulletproof Networks, says this monitoring should be accompanied by alerting. Once the monitor is tuned to avoid false positives, an appropriate person should be automatically alerted when an unusual event occurs. "SMS is absolutely perfect for that," he says.

When it comes to things such as radio frequency, monitoring is important for good wireless LAN performance, says Mark Hayes, manager of consulting and solutions at CSC. "The RF environment is not static," he says. According to Hayes, a WLAN coming online on a close neighbour's premises can affect the performance of your network.

2. Quality of service and packet shaping

One way of improving perceived performance is to ensure that the most important applications get priority. Typically, applications are allocated to classes of service (typically platinum, gold, silver, and bronze), and then policies are set for each class. For example, platinum traffic might be guaranteed at least 50 percent of the available bandwidth.

Three or four categories are typical, says Danny Price, solutions manager at Vanco Australasia, but some organisations use as many as six. A larger number is too hard to manage, he says.

Some category decisions are easy, such as blocking or severely limiting peer-to-peer file sharing, says Owen. Packeteer's software supports auto-discovery and auto-configuration, after which priorities can be fine-tuned to suit the needs of the organisation. The traffic shaping capabilities of routers are "generally all that you need to get you started," says Atkinson. "A lot of places don't take the basic steps." If further improvements are needed, the Packeteer PacketShaper is a good product, he says.

Hayes warns that people don't always understand the impact of packet shaping, which can be negative if not done correctly. "We understand the applications and how to configure the [Packeteer] devices to provide the appropriate performance for the applications [along with detailed reports that the network administrator needs]," Hayes says.

Path optimisation can be used in conjunction with service classes, says Steve Wastie, director of strategic alliances at Peribit. For example, two sites might be connected by frame relay plus a higher bandwidth VPN link via an ISP. ERP traffic might always be sent by frame relay, while internal e-mail goes across the VPN as long as the latency does not exceed 200ms. This makes good use of the infrastructure, and "is a critical enabler for us", Wastie says.

Modesto points out that you may need to shop around among providers (or get an expert to point you in the right direction) to get a WAN link with the characteristics needed for your application to work at peak performance. Price says that where multiple carriers are involved (say one in Australia, another handling international traffic and the third within the US or Europe) it's important to ensure that the different classes of service are correctly aligned for optimum performance. In particular, real-time traffic must be kept in the top class all the way through the infrastructure.

3. Compression

"You're always going to have a bandwidth limitation," says Wastie. Changes such as the perceived need for disaster recovery, ever-growing PowerPoint decks and the tension between increasingly distributed staff and increasingly centralised infrastructure soak up previously spare bandwidth, while locations in rural areas and hard-to-service facilities such as oil rigs will always have limited bandwidth.

Where this is the problem, compression could be the answer. Modern compression algorithms, including those used by Peribit and Packeteer, are able to recognise patterns in very large data streams perhaps weeks apart. This gives better results than traditional algorithms that use a limited window, perhaps as small as 1Mbit of data.

Compression is actually a combination of compression and caching, says Owen. He says Packeteer uses four different algorithms to suit the requirements of different applications. For example, file transfers can benefit from relatively slow but thorough compression, while packets for a transactional application should be handled as fast as possible.

"Having TCP rate control and the level of compression [handled by one appliance] by far provides the best value in terms of optimising the network," says Owen. The functions can work against each other if they are separated, and the most aggressive application will still win. Correctly implemented, compression can increase the throughput as much as fourfold, he says.

4. Protocol acceleration

Satellite links involve an additional round-trip latency of approximately one second, and this limits the speed of TCP/IP communication. Wastie cites a real-life example of a 1Mbit/sec line with a latency of 1.1 seconds that achieves a maximum throughput of 100Kbit/sec. TCP acceleration removes that bottleneck and allows the line to run at its nominal speed.

Adjusting packet sizes can also help, says Gibb. As mentioned above, large file transfer packets can block small packets from interactive applications. The problem is that even if the small packets are prioritised, they may be delayed for the time it takes to send a large packet. The answer is to split the large packet into smaller pieces. This can be achieved by configuring the client, server or router.

Increasing the window size so the sender doesn't wait for an acknowledgement of one packet before sending the next can reduce the effects of high latency, and incorporating error-correction information can reduce or eliminate the need for retransmission when an error does occur, explains Gibb.

5. User involvement and education

Poor performance can occur as a result of bad user behaviour, but it may be more effective to get your colleagues outside through participation and education rather than imposing harsh standards and technical lockdowns. Prichard relates a situation where a mining company in WA experienced network slowdowns at lunchtime. The cause was traced to Doom sessions between staff at the minehead and down the shaft. Once the problem was explained, play ceased. "It's

education, not Big Brother. People don't understand [the effect they can have on the network]," he says. Similarly, encouraging people to save PowerPoint files on a shared drive instead of e-mailing copies to everyone concerned can help. Hayes notes that user education may be required to discourage people from doing things like unnecessarily replicating e-mail databases from a server to their PCs.

Modesto says malware often gets inside the firewall on notebook computers, so their security is a priority and user education about safe practices is an important element of avoiding problems, in addition to locking down configurations as far as possible without excessively impinging on user activities.

HR issues can affect performance in other ways: if incentive payments to IT staff are based on technical criteria such as the uptime of WAN links, they may concentrate on these rather than business outcomes, suggests Prichard.

6. Out of band management

How often does cycling the power fix a transient problem with a server or other device? If you don't trust branch office staff with the key to the broom cupboard -- sorry, the server room -- for fear they will flip the wrong switch it can take hours to get a technician on site. Another problem is that if a device becomes misconfigured and drops off the network, you can't use the normal remote management facilities to reconfigure it.

Out of band management using products such as those from Cyclades can overcome both types of issue, and is becoming increasingly important with the trend to geographically separate data centres and systems administration staff (which may or may not include the outsourcing of administration). Charlie Waters, senior vice president for global marketing at Cyclades, says that reducing the mean time to repair a fault increases overall productivity, as well as that of the staff involved in fixing it. If a customer has 3000 servers, of which six are usually down at any one time, it is important to get failed servers back online quickly for performance reasons, even if service availability is 100 percent due to redundancy.

Out-of-band management uses separate, secure communications paths into the production infrastructure to minimise downtime. Devices such as console servers and power managers are co-located with the servers and other devices and connected to them using serial, KVM, or Ethernet links. The important points are that the connections between the administration point and these devices are completely separate from the production channels, and a single management console can support all the infrastructure components.

According to Waters, a European telco reduced overtime costs by 88 percent, the average fault fix time by 97 percent, and the total fault hours by 88 percent as a result of using this technology -- and the cost was recovered in around a year.

"There is tremendous pressure on IT managers to improve service levels and efficiency," Waters says. He says the separation of the control network from the data network is an architecture proven by the high service levels delivered by the phone system.

7. Mistimed traffic

An overnight backup process that spills into working hours can easily clog up a network. This can be reduced through user education or by taking technical measures, suggests Atkinson. For example, locking down PCs to prevent users installing software will reduce the number of files that change from one day to the next.

Backup software may respect a time window and prioritise any missed files during the next run if correctly configured. For greater flexibility, look for software that will limit itself to a certain fraction of the available bandwidth during particular hours, that way it can run at full speed during quiet times, and throttle back to a trickle feed during the working day to complete the backup as soon as possible without causing disruption. This can also be implemented through QoS features.

It's important to ensure that your hardware is fast enough for the job. Can the backup server do virus checking and compression in real time? Can it write to tape at least as fast as the data is arriving? The network isn't the bottleneck if you're using a 2Mbps link but the drive is only running at 1Mbps, Atkinson says.

Other processes can occur at the wrong time. Atkinson mentions a situation where Dell's OpenManage systems management tool had been configured to discover all devices at 10am each day, flooding the network and slowing real work to a crawl. There was nothing wrong with the software, he says, just the way it was configured.

Something similar can happen with automatic updates to antivirus and other software if too many PCs try to update at the same time. For example, the plan might be to update branch office computers primarily from a local server, with a head office server (or even the vendor's web site) as the secondary. It's easy to clog a WAN link if the branch server is down and all the PCs in the building try to update simultaneously.

"You need to be a little bit careful about the way you configure things," Hayes says, adding that organisations with international networks need to pay particular attention to timing, especially when moving bulk data between regions, as one

area's quiet time can coincide with the other's peak.

8. Citrix/thin client

Webifying enterprise applications may make for a more consistent user interface overall, but it can also degrade network performance. According to Hayes, some analysts report it can consume five times the bandwidth while delivering only one-quarter the performance.

One solution is to use Citrix-based thin client technology to reduce the amount of data flowing through the network, says Phil Osborne, senior consultant, enterprise, at Citrix Systems Australia. He says it even makes sense to run the browser on central servers -- "that's a trick we see a lot of companies doing" -- otherwise the application may run more slowly than the previous client/server architecture.

"Just don't move the traffic around the LAN or WAN unless there is a real need to do so," says Osborne. For example, large files attached to e-mails remain inside the data centre unless they are explicitly copied to a PC. He points to Flight Centre as an example, where branches have been equipped with Wyse terminals to access centralised Citrix servers over relatively low-bandwidth connections. Print traffic can put a significant load on a network in some environments, says Osborne, but the combination of Citrix's recent print drivers and products such as Exceed, Spinifex and ThinPrint reduce the traffic and increase printing speeds.

Citrix offers software that supports streaming video to a thin client, and has acquired a company with technology that will enable the use of VoIP softphones with thin clients.

It's a question of looking at the data that's being sent, and identifying a smarter way of sending it, Osborne says.

Switching to Citrix isn't the end of the story. Gibb points out that various tweaks -- such as tuning the caching of large bitmaps or the appropriate segmentation of packets or frames at the data link level -- may make an appreciable difference to overall performance.

9. Keep junk traffic off the network

Antivirus software, spam filters and firewalls all help prevent the generation of junk traffic within your network, so make sure they are enabled and kept up-to-date. Modesto says it is worth considering outsourced antispam and antivirus services, as they typically use multiple products to provide ongoing protection on the occasions when a vendor takes an extra day to provide an update for the latest virus or worm.

Atkinson also suggests blocking e-mail attachments to the extent that is feasible, and configuring software so that large attachments are held on the server as long as possible. Just because 10 people are sent copies of a multi-megabyte PowerPoint deck, that doesn't mean they are all going to open it. User education comes into this too, as it would probably have been better to store the file in a shared folder, and send a link to those 10 people. Atkinson also recommends disabling the "All" group in e-mail -- it typically comes at the top of the list, so users will accidentally select it from time to time. It's also a sitting target for mail viruses and worms. "Make patch management... and laptop security a priority," advises Modesto, though updates should be performed at night or staggered throughout the day to avoid congestion. He also warns that some popular printers run cut-down versions of old operating systems and can be affected by worms. Monitoring tools such as MRTG can reveal unexpected traffic: "a little bit of graphing goes a long way."

Users may want to install legitimate but unapproved software that adds to the load, such as utilities that load fresh wallpaper every day. A noticeable spike can occur if enough people follow suit. Or the program might hog RAM or another resource, causing poor overall performance. "It's really about knowing what's running, who's running it, and what they're doing," said Prichard.

Broadcast traffic that's not relevant to all users can also be regarded as junk. Jae-Won Lee, product marketing manager for data networking solutions at Nortel Asia Pacific, says this can be reduced by dividing the network into multiple virtual LANs (VLANs). Segregating a 100 user LAN into five VLANs will hide around 80 percent of broadcast traffic.

"For example, if an organisation has multimedia, CAD/CAM design or on-line collaboration tools that use multi-cast protocols which inherently produce a lot of broadcast traffic then these functional groups can be separated from the rest of the organisation as not to impact other traffic on the network," he says.

Although it's important to monitor the network, Atkinson warns that it is possible to overdo things by sending too many pings and test frames. Some of his customers were losing one third of their bandwidth to multiple and inappropriately configured network management tools until he set them straight.

10. Has your network kept up with any changes?

A network can be perfectly designed and implemented to a specification, but requirements change. New applications are added, traffic patterns change, staff are moved between locations and so on. Nortel's Lee points out that older LANs were

often designed with an aggregation layer between the wiring closets serving individual floors and centralised resources, reflecting the use of physically distributed departmental servers and other workgroup infrastructure. It also reduced the number of ports required on the core switches.

The consolidation trend seen over the last few years means that the majority of traffic now flows from desktop PCs to central servers, so removing the aggregation layer will improve performance. This may mean increasing the number of ports on the core switches, but the improvement will be especially noticeable with voice traffic, Lee says.

Hayes says that too often, those deploying an application do not consider the effect it will have on the network, while those responsible for the network do not always understand the effect changes will have on applications. The placement of servers should be optimised in terms of network resources, cost and performance. For example, it may make sense to move an application server closer to the users -- but what effect will that have on communication between the app server and the database? It might be better to move to a thin client architecture, or to rearchitect the entire application, he suggests.

Similarly, the use of spanning tree protocols to handle redundant network links is no longer appropriate, says Lee. Not only does it require the "backup" link to sit idly in reserve, but it also takes between eight and 50 seconds for individual sessions to reconverge on the other link following a failure. That is no great drama for most applications, but it is hopeless for VoIP traffic. Nortel's Split Multilink Trunking (SMLT), an extension of the 80213ad standard, enables simultaneous use of both links and has a reconvergence time of less than one second, he says.

According to Roland Chia, national business manager at Dimension Data, IEEE 802.1d Spanning Tree eliminates network loops in a LAN switching environment but can cause network instability if not configured correctly, for example when a misconfigured switch with highest priority is connected to a production network. "Best practice is to configure the LAN with Layer 3 switching or use Cisco proprietary advanced features such as Spanning Tree Rootguard feature," he says. Hayes says network architecture is about having the right devices in the right places doing the right things for the job, so if you've got a Layer 3 switch at the core of the network, use it as a Layer 3 switch.

Adding VoIP represents a major change. David Paddon, managing director of NSC Enterprise points out that if there is a delay of 30 seconds in transferring a spreadsheet from one place to another, with VoIP its integrity is still intact. That's not true for voice or video, where all packets must arrive in a timely manner.

Think about power outages too -- people expect to be able to use their phone during a blackout. This requires power over Ethernet (PoE) to the handsets, plus backup power to the entire network, Paddon says. People have a "five nines expectation of performance" from a phone system, says Hayes, who also recommends redundant, dual-homed floor switches to ensure high uptime. CSC has installed such a system at its Australian headquarters in Sydney. A high-availability LAN is supported by PoE, UPS and a generator in case of prolonged outages, along with dual links to the data centre using diverse paths and infrastructure. "We see voice as being the most critical application on the network," he says.

Atkinson warns that software configurations need to reflect network changes. One organisation had used frame relay to connect its head office, state offices and branches in a hierarchical arrangement, and updated files were sent to the state offices and then onto the branches. That worked well until it switched to a DSL network with a star topology: each time a state office sent an update to a branch, it went via head office. The new arrangement was "four times as fast, but twice as slow," says Atkinson, but the problem was overcome by having the updates sent directly from head office to each branch.

When adding switches or servers to a network, you should not rely on automatic Ethernet configuration, warns Chia. "Automatic configuration between vendors is not standardised and should always be manually configured to match," he says. Hayes agrees, saying full or half duplex settings should always be explicitly configured to match.

Case study

The Department of Employment and Workplace Relations (DEWR) uses several of the techniques described in this feature to get the best performance from its network.

Among its other functions, DEWR provides IT services to the Indigenous Coordination Centres (ICCs). These centres were previously regional and state offices of the Aboriginal and Torres Strait Islander Commission (ATSIC) and Aboriginal and Torres Strait Islander Services (ATSIS).

Ian Rowe, director of communications and IT security at DEWR, says a Citrix thin client arrangement was adopted for performance and operational reasons as some of the ICCs are in remote locations. He says Citrix delivers better performance across the WAN, and it is much easier to maintain centralised servers. The data centre also provides better environmental control and physical security than is available at remote sites.

There were initially some performance issues, such as a noticeable lag between pressing a key and the character

appearing on the screen. This was overcome by using the Network-Based Application Recognition feature of the Cisco routers to give Citrix traffic top priority. This arrangement was fine-tuned using a Packet Description Language Module to assign the highest priority to Citrix KVM (keyboard, video, mouse) traffic along with real-time video streams. Conversely, Citrix printing packets (for example) are given a low priority. "That's been very successful for us," says Rowe. Some user retraining has also been required, such as the teaching that opening a file via Internet Explorer is a lot quicker than doing so through My Computer.

DEWR also gives backup traffic a very low priority to avoid impacting normal operations in the event that it is not completed before the start of the business day. It typically gets 100 percent of the bandwidth at night when there is little other network activity.

On-demand video is cached by content engines at each location, and links to the files are automatically redirected to the local copy rather than going across the WAN. Any updates are given very low priority, just like the backup operations.

"Using PDLM and NBAR has been a real breakthrough for us [in terms of getting good performance with Citrix]," says Rowe. DEWR chose not to use a packet-shaping appliance because it wants to keep the network as simple as possible and wanted to avoid any extra latency, he explains. "If we can do something in the router, our preference is to do it there."

Various measures are taken to keep unwanted traffic off the network. The routers only propagate TCP traffic, isolating any other protocols to the local network where they originate.

Anti-virus software is installed on all servers and desktops, and e-mail is scanned at the gateway, on the Exchange server, and on the desktop. Three different products are used to reduce the risk of a new virus slipping through all three layers. SpamAssassin is used to flag rather than delete spam. Rowe plans to augment this by activating the relevant features of Exchange and Outlook, but says it would be better if spam was filtered at the ISP level, before it reaches the department's network at all. Sometimes malware does get through. DEWR was affected by Welchia, which generates a lot of network traffic. Rowe says this activity was picked up by an IDS and as a temporary measure the Welchia traffic was routed into a black hole.

Executive summary

- Use appropriate monitoring and analysis tools to understand what's happening on your network.
- Quality of service and packet shaping tools and features can prioritise the most important traffic.
- Out-of-band management can improve overall performance through greater uptime.
- Protocol acceleration may relieve the effects of high latency.
- Compression may relieve bandwidth constraints, but should be combined with traffic shaping.
- User involvement can be more effective in stopping undesirable practices than a punitive approach.
- Ensure backup, update and network management traffic does not impinge on "real" applications due to inappropriate configuration.

This article was first published in Technology & Business magazine. [Click here for subscription information.](#)