



# A Smorgasbord of Small Topics

Peter J. Welcher

## Introduction

I've been saving up some miscellaneous topics that I thought might be interesting. They are:

- 1 Deploying QoS in the Real World
- 1 Hidden Connectivity Failures
- 1 Things Application Developers Ought to Know (But Apparently Don't)
- 1 What Do You Do When You Outgrow Your T1

Let's dive right in.

## Deploying QoS in the Real World

I've encountered a couple of situations lately where folks were perhaps a bit surprised at how complex Cisco QoS can be. (And thanks to them for sharing their thoughts!) I too end up being surprised occasionally, despite having been doing Cisco-based QoS for over 6 years now. I do find myself wishing the complexity level would go down some, especially at the configuration level. But the world of QoS is still evolving.

One comment about apparent complexity. Quite some time ago the Cisco documentation guides lost the distinction between features and important features. As we all know, 90% of the Cisco commands are "knerd knobs" that are only occasionally needed. The Config Guide unfortunately isn't organized -- prioritized -- around that fact. I'd somewhat like to see each area have sections titled Essentials, Useful Features, and Tweaking Your Configuration (or something like that). Important, Somewhat Important, Knerd Knobs.

This is particularly true for QoS. I keep seeing people fiddling with COS and DSCP maps. The defaults are usually fine, and this is minor. The same applies to assigning COS values to queues, although in one or two cases, the defaults are NOT fine and you have to do that. Otherwise, leave the knerd knobs alone!

With the 6500 series, you do have to be careful. With most of the LAN line cards, you have a MQC interface that only lets you do classification and marking or policing. You can put in priority and bandwidth commands, you can apply your QoS policy to an interface -- but it will NOT do anything! The words "Serious User Interface bug" do come to mind. By the way, shaping is also not available for most LAN cards in the big Cisco switches. You can get it right now, but as far as I know, the list price for a blade supporting it is up around \$60K (better than the \$110K it used to be).

Another gotcha I've seen in deployed devices is accidentally omitted the "mls qos" or "qos" command. If you don't enable QoS, all those other lovely commands don't do a darn thing. Oops! Lesson learned: check things are actually getting marked. I suggest doing packet capture using Ethereal or using show commands. Unfortunately, the QoS-related show commands in most of the switches are mostly rather useless, as far as I'm concerned. In testing, put in an absurdly low value for bandwidth or priority and see if things get dropped. Test your configuration. And then be very careful deploying. And be suspicious -- verify your work, since it is very easy to be fooled into thinking QoS is doing something, when it isn't.

Another thing that seems to surprise people is the variety between platforms, even within Cisco. Well, that's the Cisco development process. Different products have different management teams and engineers. Cisco's overall approach is to encourage initiative. My understanding is that otherwise you lose many months of time-to-market while product and marketing folks coordinate features, etc. (I am trying hard to NOT mention IBM and HP here!)

Some folks I've enjoyed working with had some recent surprises deploying QoS in a mix of routers in a large government WAN. (Thanks, Dean!) It turns out the various cards, such as the ATM port adapters in 7200's, may or may not support certain QoS features. One recent IOS version dropped support for VC-bundles on ATM. The surprise was compounded by the fact that the Software Compatibility tool on cisco.com did not catch the issues -- it just wasn't fine-grained enough. (Understandably a hard problem for the Cisco people supporting it -- but frustrating in this case). Ouch!

What advice do I have about this latter? Test everything you can in the lab, and even then be prepared for surprises if there are hardware variations in the field. We had tested in the lab. You just can't test to that level of detail, it would take prohibitive amounts of time.

If you're looking for Best Practices, be aware they are very situational (just like Security policies). It all depends on what traffic YOU think is important. There is no one best answer that fits all situations. If you're developing a product that interacts with QoS somehow, you'd better be prepared to deal with this. If you're going to tell people what to put into their Cisco routers and switches, you need to deal with at least the top-selling models, and be prepared to update the list and your advice as you encounter new models or quirks. If you're selling a product that configures the routers for your customers, that's part of the value you can add to the product.

I recently did some work that lead to heavily reviewing the (sparse) Cisco RSVP documentation. The documentation really doesn't say much about how RSVP actually works in the Cisco routers. In particular, the documentation does say RSVP works with weights with WFQ. Fine, I believe that, and I know enough about that to believe it probably does what is needed. Does RSVP also do that with CBWFQ/MQ? Not documented, as far as I could see. The RSVP LLQ feature is mis-titled, it is really Priority Queueing with WFQ, a far different thing. Etc. I ended up with the feeling that the only thing Cisco RSVP seemed guaranteed to work with was WFQ.

Concerning CBWFQ and other settings I really wanted to test almost everything relating to RSVP in the lab, to see what the routers actually do. That's a shame, since the proxy gateway products have been extended, providing the opportunity to use RSVP to control call and video traffic in your WAN -- an idea I really like. That's because H.323 zones are messy at any scale, and are also high-maintenance. RSVP adapts better to managing bandwidth when you want to take link failure/failover into account. Of course, vendor support for RSVP is yet another question to consider if you're thinking about using RSVP in a design.

I trust you won't confuse me with Forrest Gump when I say "and that's all I have to say about that". (I have lots more to say, but it either wouldn't be printable, or would contain the sort of detail appropriate for consulting but not for an article like this). I can suggest consulting the Cisco documentation heavily, also the SRND Guides for QoS (System Reference Network Design Guides). Good stuff! (<http://www.cisco.com/go/srnd>).

## Hidden Connectivity Failures

Several times recently I've seen situations where a link fails but the routers (and operations staff) don't know it. One revolves around copper-to-fiber media converters. If they don't support some form of fiber cut/one way detection and convert that to copper link signal, you can get some interesting failure modes. Like neither end noticing the link is down, no Spanning Tree reconvergence, just black-holing packets.

This is also a concern for customers of Metro LAN services. Your link to a switch or optical access device may well stay up, even if the vendor SONET / DWDM / Spanning Tree is having a bad day. In which case your edge device happily forwards packets or frames into the void -- more black-holing!

We've also had some requests for assistance from enterprises that (wisely!) wanted to use two MPLS VPN providers, and failover between them when one was having connectivity issues. Their inexpensive providers were doing static routing for them.

This turned out to be fairly interesting to think about, and fit in nicely with my upcoming MPLScon presentation (which will be posted shortly after it is presented). Last year's version is already online at <http://www.netcraftsmen.net/welcher/seminars/mplscon05-buyersguide.pdf>.

The short answer is that you either need to be managing the Customer Edge (CE) routers yourself, or find a common routing protocol your MPLS Service Providers (SP's) will provide to you (good luck on that!). I thought of GLBP but I don't think it helps sufficiently. Loss of edge connectivity from PE to CE router is one possibility. The other intriguing issue with MPLS is a routing issue and loss of connectivity "in the cloud". This could result in your continuing to receive routes from (or use static routes to) the affected provider, in which case you would forward and ... black-hole packets. In most MPLS scenarios, that shouldn't happen. But if your deal with the MPLS SP uses static routes, then this would be a very real problem.

There seems to be a recurring theme here!

What can you do to detect black-holing of packets? The classic answer is to run something with a heartbeat or keepalive. The routing protocols EIGRP and OSPF do that with Hellos. GRE tunnels do it. BGP does it, indirectly. If those don't work for you, IP SLA (former SAA) and object tracking might fit in with detecting black holing. Having just written about Optimized Edge Routing (OER), that too came to mind.

So part of my MPLS VPN consumer advice is: if you're getting MPLS VPN WAN services, especially from two providers, you might really want to look into dynamic routing, hopefully where both providers will provide EIGRP or OSPF to you, EBGP if they must.

If you've got static routing (which we are seeing more and more of, smaller / cheaper MPLS SP's apparently assume they're the only SP your network would ever be using) then you need to revisit that list above. Which of these approaches you pick is very situationally dependent. I sure would want to keep things simple. In one case, we considered using the MPLS VPN providers' static routes (the contract was already signed) to route between the CE routers attached to each MPLS SP. That means you could bring up an EBGP session, say using a different BGP AS for each site. Run IBGP between the two CE routers at each site. If there were only 1 router at each site, connected to both MPLS SP's, you'd have to control route advertisement loops instead. The other idea that comes to mind is trying to do some form of EIGRP multi-hop using neighbor statements. Messy, but no worse than setting up lots of EBGP. Faster converging than EBGP. Would that be a Good Thing or a Bad Thing?

## Things Application Developers Ought to Know (But Apparently Don't)

This one risks turning into a rant, so I'll keep it short.

At birth, every future software developer should be issued a WAN simulator or pair of old Cisco 2500 routers and taught the bandwidth command. They can compile code, etc. on a fast modern LAN, but should be forced to do their program testing over a slow WAN link, preferably one with a non-zero Bit Error Rate as well.

You'll instantly know why I say this, when you first encounter some relational database program (or other program) that runs fine for the developer but is slow as a snail on the WAN. It is **always** the network's fault, at least in terms of blame. It rarely is the network's fault, in actuality. Many sites buy OPNET ACE to prove that no, it really isn't the network. The problem is that by that point, it is far too late to recode the application, which may well have fundamental architectural problems. After the blame-storm dies down, you're left with the fact that way too much money and time has been spent, the program may not be fixable -- and the pressure comes back on the network team to somehow make the problem go away. This is where some clear simple slides about bandwidth and delay versus throughput are probably appropriate.

I don't usually need to access any deep knowledge here. Ethereal could be all I may need to see the problem. If a program moves massive amounts of data between server and client, it probably isn't going to be WAN-friendly. And if it does a lot of back-and-forth ("ping-ponging") with small packets, it probably isn't going to like the WAN.

Good to great application developers and database analysts are highly aware of this, and do a good job. But there are enough of the other kind around to make life very interesting.

By the way, application developers also ought to know that broadcast is not a good choice of destination address. Either learn and use multicast, or figure out a different approach. We recently saw a situation where point of sale application broadcasting clobbered the WLAN badly enough so the point of sale became point of NO sale. Oops!

Independent of the question of programmer skills, we can look out for protocols and applications that bog down on a WAN. Situations where this comes up:

- 1 Database tools that do pattern-matching on the client (Citrix is usually the way to fix this)
- 1 Database tools or report tools that pull records down sequentially rather than say 20 at a time
- 1 Microsoft File Shares on the WAN, especially via IPsec over the Internet
- 1 Using NAS instead of SAN for storage

In case you're wondering about the last two, they are variants of the same thing. When you access a file system (Microsoft server or NAS), the code recurses down directory trees. At each stage, it has to check permissions then open the next level. So going to a directory that is 5 folders deep may take 10 or more packets in each direction. If you are

dragging and dropping many files, that gets repeated for each and every file. Multiple that by the round trip delay on the WAN and you'll come up with ... SLOW! By way of contrast, a SAN device does (virtual) block level i/o. In effect: slam, there goes the data, write to disk, acknowledgement back, done. Much less back-and-forth (ping-ponging). Hence faster on the WAN (or even LAN)! I just happen to have read a Cisco Networkers presentation about their caching Wide Area File System. It states that MS Word does a lot of synchronous back and forth (as in, 1 MB file leads to 1000 packets), so even with a low RTT of 40 msec that's 40 seconds to save the file.

You might also want to think about managing your managers' expectations. This has come up in discussions with a customer. What you don't want is for the expectation to somehow creep in that you will be doing fast Microsoft file shares or VoIP / IPT over your low cost best effort WAN connections. If you have gone to IPsec over the Internet, you might think about repeatedly explaining to management why it is so cheap, and what you're giving up. Namely, reliability, QoS, fast response if you take packet drops, management access to routers in the path, etc. That's part of what MPLS VPN carrier SLAs are supposedly buying you. Even there, management access is needed to verify that you are in fact receiving the SLAs you are paying for. As the saying goes, you get at most two out of {cheap, good, fast}.

## What Do You Do When You Outgrow Your T1

We are living in curious times. I'm seeing situations where folks upgrade from a fractional-T1 Frame Relay circuit to an older-style T1 leased line, because T1's have gotten so cheap. Others are replacing the fractional-T1 access circuits with T1's, because doing that is cheaper (and faster). There are regional variations and all sorts of local oddities, where there seems to be a lot less commonality across markets than there was for the last 10 or so years.

One thing I do see is a growing WAN divide. Sites in or near big cities are finally getting to tap into fiber glut, and seeing dark fiber or Metro Ethernet services at good prices. Sites in rural locations may only be able to get still-costly T1's, or even more costly microwave-based T3's. Cable providers are deploying fiber to sites in some rural locations. In the really rural locations, I suspect copper is all you're going to see for quite some time.

If you're at sub-T1 speeds, T1 is the no-brainer next step up.

If you're already at T1, what do you do? T3's can be awfully expensive. Even if the recurring costs are low, the interface and connection costs are fairly high. What are the alternatives?

One alternative is to get multiple T1's. You might then allow routing to load-share. You could also run Multilink PPP with fragmentation turned off, say on up to 4-8 T1's.

Another alternative is now FR bundling, FRF.16. Multilink Frame Relay is supported in Cisco gear, and by some FR providers, and has been for a while now. See for example [http://www.cisco.com/en/US/partner/products/sw/iosswrel/ps1829/products\\_feature\\_guide09186a0080087079.html](http://www.cisco.com/en/US/partner/products/sw/iosswrel/ps1829/products_feature_guide09186a0080087079.html).

If you're doing FR, the first place that gets tight on bandwidth is usually Headquarters (HQ). Many carriers trunks used to only support FR up to DS3 speeds. That's where FR to ATM service interworking is useful. You get the carrier to deliver packets to HQ on an ATM circuit. That opens the door to OC-3 or OC-12 connections. The hardware isn't cheap, but you're not widely deploying it either. It seems like by the time OC-48 rolls around lately, folks have found a way to do some form of Ethernet to the carrier. FR or ATM to Ethernet service interworking is something to look forward to (and try to obtain if you need it).

If you're a fan of ATM, you can always do Inverse Multiplexing. Doing this with up to 4 T1's is fairly cost effective at the access layer. Just be careful in planning to do IMA, because if your bandwidth growth rate is fast enough, you'll have to amortize the interfaces (and possibly the routers) over a 2-3 year period.

I'll note in passing that we're seeing peoples' networks adding bandwidth at a rate of doubling the speed every 2 years or so, sometimes faster. So make sure you take growth into account, so you don't have to go through Yet Another WAN Upgrade 2 years later -- unless you like the travel, or figure it is a form of job security.

## Summary

Thanks to our David Yarashus for suggesting the topic about outgrowing your T1. Obviously I like it!

I have no separate reference links this month.

If you have insights or war stories on any of the above, I'd be interested in hearing them. My email address is below.

Your comments, questions, and suggestions for future articles are of course welcome! See below to decipher Pete's email address.

---

Dr. Peter J. Welcher (CCIE #1773, CCSI #94014, CCIP) is a Senior Consultant with Chesapeake NetCraftsmen. NetCraftsmen is a high-end consulting firm and **Cisco Premier Partner** with multiple specializations, dedicated to quality consulting and knowledge transfer. NetCraftsmen has eight CCIE's, with expertise including large network high-availability routing/switching and design, VoIP, QoS, MPLS, IPSec VPN, wireless LAN and bridging, network management, security, IP multicast, and other areas. See <http://www.netcraftsmen.net> for more information about NetCraftsmen. Pete's links start at <http://www.netcraftsmen.net/welcher>. New articles will be posted under the Articles link. Questions, suggestions for articles, etc. can be sent to **[pjw <at> netcraftsmen <dot> net](mailto:pjw@netcraftsmen.net)** (formatted this way to fool email harvesting software).

5/7/2006

Copyright (C) 2006 Peter J. Welcher