Dragon Slayer Consulting

Eradicating The VMware WAN Performance Barrier

Overcoming Data Packet Loss, Latency, and TCP/IP Packet Loss Management

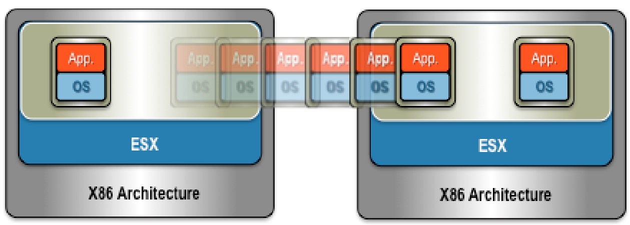
Marc Staimer, President & CDS of Dragon Slayer Consulting
marcstaimer@comcast.net 503-579-3763

Introduction

VMware is an incredible tool. It seems to offer a lot of answers to many server issues. Issues such as application availability, simplified disaster recovery, reduced infrastructure (and cost), simplified IT administration, and "greener" computing.

The key to VMware's value is its ability to measurably increase hardware utilization, provide greater application deployment flexibility, while facilitating highly cost effective server consolidation. This is very compelling, which makes it easy to see why VMware has captured more than 50% of the market.

VMware has much touted industry leading capabilities including Vmotion (transparent live migration of applications without a disruption); Storage Vmotion (live migration of storage volumes without a disruption); Site Recovery Manager – SRM (near instantaneous recovery of down machines at a local or remote site); Distributed Resource Scheduler – DRS (on-demand allocation of hardware resources based on QoS policies per virtualized guest); and now Continuous Availability (non-stop applications even if there is a hardware failure).



These are gripping capabilities that keep VMware a step ahead in the server virtualization market. So why then is there so much market frustration with VMware wide area network performance?

It's because not surprisingly, VMware does nothing to fix well-known WAN throughput issues (the same is true of all other hypervisors as well). In fact, server virtualization tends to exacerbate those WAN throughput issues because of increased WAN traffic from/to each server.

The VMware Wide Area Network Problem

Wide area networks rarely provide the effective bandwidth throughput that users expect. This is analogous to the difference between raw storage and usable storage. The amount of usable storage is always considerably less than the amount of raw storage because of file system, RAID, and data protection overhead. The amount of actual effective bandwidth throughput is also less than the "Raw" or rated bandwidth. Why the actual effective bandwidth throughput is so much less than rated bandwidth is considerably different from why usable storage is so much less than raw storage. The bandwidth differences are the direct result of the amount of data packet loss, end-to-end (or round trip) latency, and how each is managed.

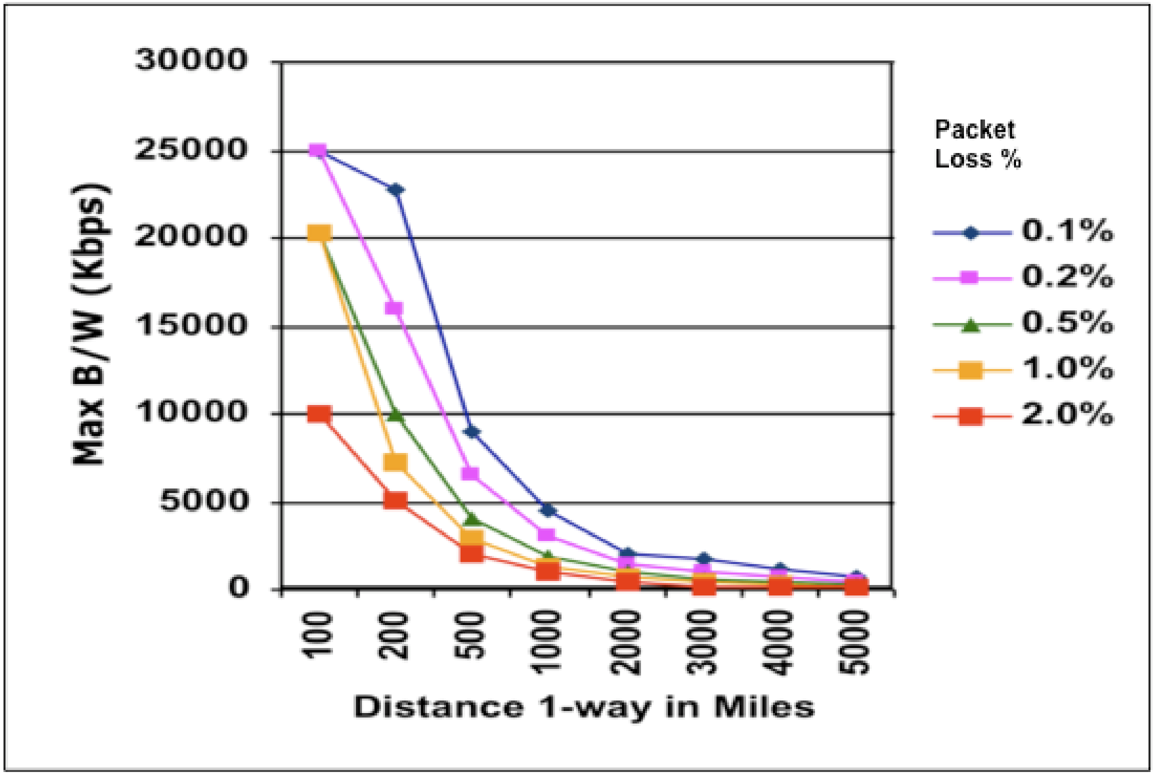
○ *Data packet Loss, Latency, and how it's handled by TCP/IP*

Data packet loss is unavoidable on a WAN and can be devastating to effective bandwidth throughput. It closely correlates to jitter, BER (bit error rate), and congestion (buffer router overflows), on the WAN. Unfortunately, TCP/IP does not handle packet loss on the WAN very well.

When a data packet is lost, standard TCP/IP requires that lost data packet and every data packet that was transmitted behind it, to be retransmitted. This is because TCP/IP is stream orientated. On a LAN this is not a big deal since the round trip time (latency) between the sender and receiver is very small. It's a very different story on the WAN.

There are two types of round trip latency or delay. The first is speed of light latency which is the amount of time it takes for light to travel roundtrip from the sender to the receiver and back. The second is TCP latency, which is the roundtrip delay penalty, introduced by TCP processing. High network latency means a lot more data packets are transmitted after a lost data packet and before that loss is detected. The result is every one of those data packets has to be retransmitted. But it gets worse.

TCP/IP cuts its transmission speed in half each time a data packet is lost and slowly ramps up to maximum speed. If another data packet is lost before the ramp up is complete, the speed is cut further. The reason TCP was designed this way is primarily because TCP assumes data loss is due to congestion and so it minimizes congestion opportunities. That works great on a high-speed LAN with very short distances. On a WAN with high latency and data packet loss, it does not work so well.



When jumbo frames and dynamic window sizing is utilized across a WAN with data packet loss, the amount of data requiring retransmission is even greater. The net result is even lower effective bandwidth throughput.

Typical data packet loss on most WAN links averages approximately .1%. However, it can run as high as 6% or more. Even .1% will decimate effective bandwidth throughput.

The overall effect of data packet loss, WAN latency, and TCP/IP data packet loss management is in a word, mind-boggling. A typical DS3 (45Mbps) WAN can be (and often is) reduced to less than 5Mbps of effective bandwidth throughput (actual end user situation). An OC3 (155Mbps) can be reduced as low as 7Mbps (actual end user situation). The rule of thumb is that most users will see somewhere between 5 and 30% effective bandwidth throughput.

So then how does VMware exacerbate effective bandwidth throughput? Many of the VMware consolidated servers are either being moved from the ROBO (remote office branch office) to the data center, or the ROBO VMware servers are integrated with virtual networked storage that's replicated back to the data center.

- ***ROBO VMware Server Consolidation WAN Quandary***

Moving the consolidated ROBO servers to the data center moves the data and applications away from the actual users, and places it across the WAN. The good news is all those VMware consolidated servers, applications, and data are now simpler to manage, protect, and administer because they are collocated with the IT talent. The bad news is that WAN data packet loss, latency, and TCP/IP data packet loss management will most likely cause user response times to go up by orders of magnitude while productivity and ROBO user morale declines in equal proportions. This is not a good formula.

❍ *ROBO VMware Server Integration with Replicated Networked Storage*

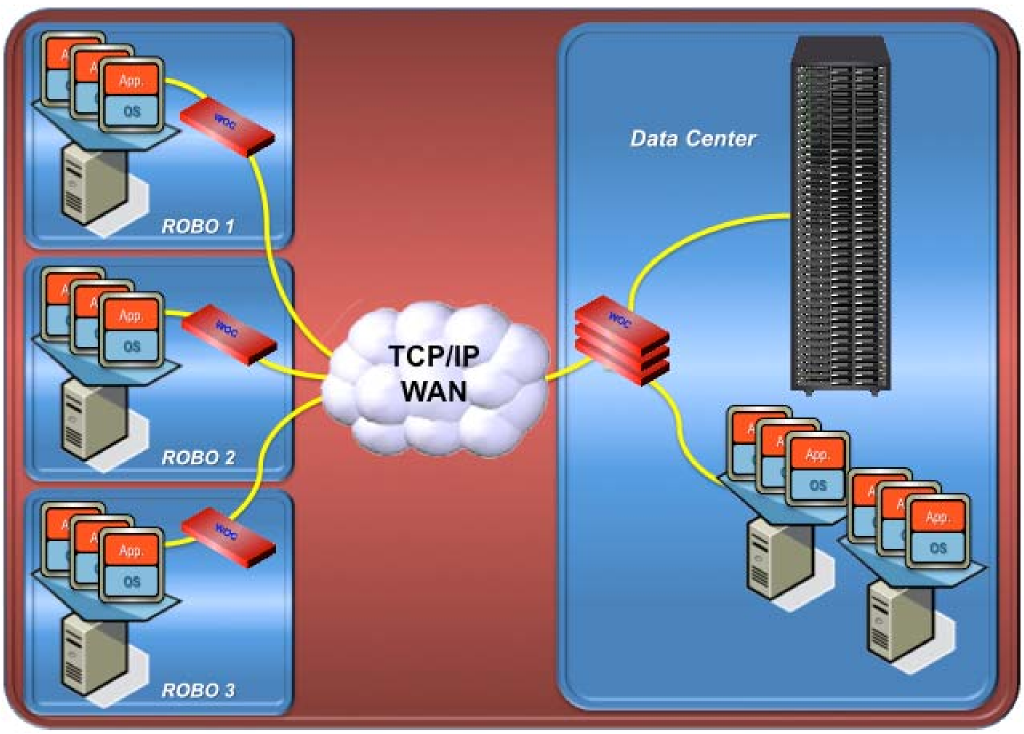
VMware requires networked storage to utilize most of its advanced capabilities. And it also requires that the storage be replicated locally or remotely over the WAN for any type of DR (such as Site Recovery Manager).

Implementing networked storage at the ROBOs can be prohibitively expensive (especially if the storage is Fibre Channel). Additionally, the ROBO IT storage skill sets may be lacking. This usually means the ROBO VMware storage is either NAS or iSCSI SAN storage. But even that can be too expensive since it requires separate storage and infrastructure to manage. To reduce the networked storage costs, virtual networked storage and virtual SANs that run as software on VMware guests (a.k.a. virtual storage appliances) have been developed. Virtual SAN appliances are available from FalconStor Software, LeftHand Networks, Seanodes, DataCore Software, and a virtual NAS appliance available from Trellis. More are sure to follow.

Regardless of whether the networked storage is externally or internally integrated with VMware, it has to be replicated over the WAN. Replication requires moving a lot a data. And once again the problems with WAN data packet loss, latency, and TCP/IP data packet loss management will rear its ugly head. WAN data packet loss will make effective bandwidth throughput too low to provide the level of data storage replication to meet data protection and DR requirements.

Conventional Wisdom Work Arounds And Their Pitfalls

The conventional wisdom to solve the WAN effective bandwidth throughput problem is to implement WAN optimization controllers or appliances (WOCs). These appliances are primarily special purpose x86



servers containing proprietary WAN acceleration software. The WOCs are deployed in the data center and at each ROBO while working together to intercept and accelerate traffic flows between ROBO users and the data center servers. Most WOCs are designed to perform some or all of the following operations:

- Data compression and/or de-duplication
    - So that less data is required to traverse the WAN link.
- Data caching
    - Latest copy of files and or data cached on local ROBO storage so that the WAN link does not always have to be traversed.
- Protocol optimization

- Application specific protocol termination at each end (spoofing) and command stacking. Works well with HTTP, HTTPS, CIFS, and to a lesser extent NFS.
- Forward Error Correction (FEC)
  - Additional error correcting data packets are added to a group of data packets that enable the receiving end to a limited extent to regenerate a lost packet in the group.
- Traffic prioritization or shaping
  - WAN traffic is separated into different categories, such as business critical applications, general business applications, background activities, and personal use to maximize throughput based on priority.

Properly implemented, WOCs can and do mitigate the WAN effective bandwidth problem for specific types of data. Most WOCs also have serious limitations and can be prohibitively expensive.

WOC Pitfalls

WOCs for the most part tend to work reasonably well with small packet size and relatively small amounts of data. They do not work nearly as well with the large amounts of data that is part of storage replication from the ROBO to the data center and especially between data centers.

WOCs performance enhancement is seriously diminished as the bandwidth increases more than 10Mbps (which is becoming far more common in this age of cable access.)

Many of the WAN optimization techniques utilized requires ROBO IT skills that may not be there. Some techniques are application specific and often will not apply.

WOCs are closed servers. This means upgrades are vendor controlled and will lag x86 server advances while requiring expensive user disruptive forklift upgrades.

Start-up and operational costs can be as much or more than consolidating the servers at the ROBO and not back at the data center.

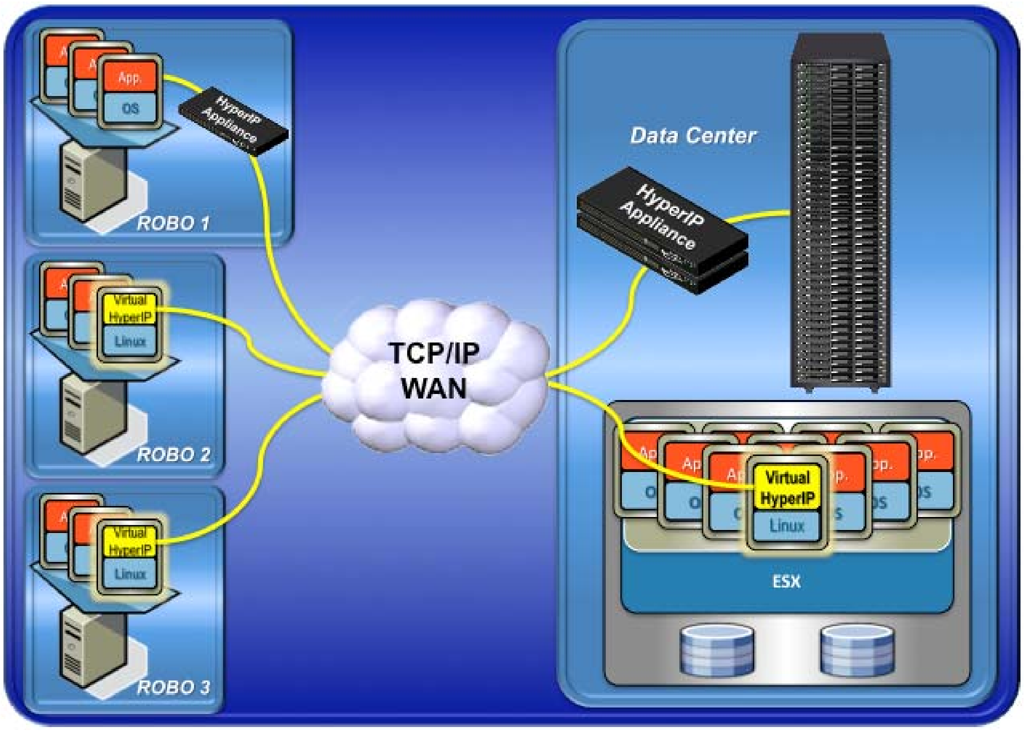
There has to be a better way to eradicate the VMware WAN effective bandwidth throughput issue. There is.

The NetEx VMware WAN Performance Answer: Virtual HyperIP

Virtual HyperIP is a data transport optimizer as compared to most WOCs that primarily accelerate specific applications. Essentially Virtual HyperIP significantly reduces bandwidth for mission critical requirements, freeing up bandwidth for chatty applications, such as those using CIFS, NFS and HTTP. Virtual HyperIP is software that operates on the VMware ESX as a guest while alleviating the effects of data packet loss, latency, and TCP/IP data packet loss management to the point where it becomes unnoticeable. Virtual HyperIP is also compatible with HyperIP appliances as well. More importantly, it does not come with the pitfalls of WOCs and in fact has significant advantages including:

- Complete leveraging of the VMware virtual server environment.
  - Meaning that Virtual HyperIP works with Vmotion, SRM, Continuous Availability, and even DRS, which allow it to use resources as required, based on policy prioritization.
  - Virtual HyperIP doesn't require stand-alone WOC hardware or appliances at the ROBOs or the data center.
  - Virtual HyperIP is backward compatible with HyperIP appliances and can work with them in a network.
  - And allows many-to-many or many-to-one network configurations.
  - And even though Virtual HyperIP does not require WOC hardware at the data center or ROBO, it will work with locations that are running HyperIP in a WOC.
- Retransmissions of ONLY the lost data packets.
  - Virtual HyperIP is a block orientated protocol, and assumes packet loss is due to network errors and uses receiver feedback to determine if congestion is an issue.
  - Virtual HyperIP reorders the retransmitted lost data packets at the receiving end.
  - Completely mitigating or eliminating the effects of data packet loss up to 6% per WAN for the applications.
- Termination of TCP/IP at each end eliminating TCP latency.
  - Mitigating or eliminating performance degradation as a result of distance latency up to 46,000 circuit miles.
- Block level compression from 2:1 to 10:1 depending on the data's compressibility and prior applied deduplication.

  - Effecting a dramatic reduction in the amount of data that moves across the WAN.
- Concurrent multiple TCP data stream aggregation from dozens to hundreds of applications.
  - Maximizing the overall throughput.
- Rate-limiting from 1 Mbps to OC12 speeds including (by time of day)
  - Making Virtual HyperIP a good considerate citizen on the WAN.
- Application transparency.
  - Virtual HyperIP will work with ALL TCP/IP apps and is not app protocol dependent.
  - Ideal for the large data applications such as networked storage replication, backup, and server-to-server replication.
- Bandwidth scaling up to 1Gbps.
  - Provides a unique Synchronization and Recovery on Demand feature set which allows for expanded performance, free of charge, for a period of 10 working days.
  - Virtual HyperIP performance enhancements do not degrade as speeds increase.
  - No forklift disruptive upgrades, just a simple matter of license keys.
- Incredibly cost effective.
  - Virtual HyperIP is licensed to the bandwidth required and upgraded as needed.
  - It's always more cost effective than adding more bandwidth, which rarely if ever solves the problem.



The end result is that Virtual HyperIP provides 80-90% utilization of the available bandwidth between data centers and/or ROBOs up to OC12 rates versus the 5 to 30% utilization that is common. Applications will see up to 10 times the performance and can see an effective bandwidth throughput that can be actually ***greater*** than the rated throughput (depends on the effective compression ratios).

Conclusion

VMware is a great boon to server consolidation and application availability. The biggest barrier to implementation at ROBOs is the effective bandwidth throughput over the WAN. Standard WOCs only sometimes solve the problem and often at a prohibitive cost.

Virtual HyperIP is the first solution to truly leverage the value of VMware. It provides a complete answer to increasing the WAN effective bandwidth throughput to where it needs to be at a rational cost.