VMware vCenter™ Update Manager 5.0 Performance and Best Practices

Performance Study

TECHNICAL WHITE PAPER

Table of Contents

Introduction....3
Benchmarking Methodology....3
Experimental Setup....3
vSphere Update Manager and vCenter Server....3
ESX/ESXi System....4
Virtual Machine Operating Systems....4
Network Simulation Software....4
Network Configurations....4
Update Manager Server Host Deployment....5
Performance Tips....6
Latency Overview....7
Performance Tips....8
Resource Consumption and Job Throttling....9
Optimization for Cluster Remediation....10
Performance Tips....11
Host Operations in WAN Environments....11
Performance Tips....12
Network Throttling for Host Staging and Remediation....12
Performance Tips....15
Conclusion....15
References....15
About the Authors....16
Acknowledgements....16

Introduction

VMware vCenter™ Update Manager (also known as VUM) provides a patch management framework for VMware vSphere®. IT administrators can use it to patch and upgrade VMware ESX® and VMware ESXi™ hosts, upgrade VMware Tools and virtual hardware for virtual machines, as well as upgrade virtual appliances.

As datacenters get bigger, performance implications become more important for patch management. This study covers the following topics:

- Benchmarking Methodology
- Update Manager Server Host Deployment
- Latency Overview
- Resource Consumption Matrix
- Cluster Optimization
- Host Operations in WAN Environments
- Network Throttling

Benchmarking Methodology

Experimental Setup

VMware vCenter Update Manager 5.0, VMware vCenter Server 5.0, ESX 4.1 and ESXi 5.0 were used for performance measurements. WANem 1.2 was used for simulating a high-latency network with packet drops. Microsoft Windows XP SP2 was used for virtual machine hardware and tools scan and remediation.

vSphere Update Manager and vCenter Server

- Host Computer: Dell PowerEdge 2970
- CPUs: Two 2GHz AMD Opteron 2212 dual-core processors
- RAM: 16GB
- Hard drives: Eight 73GB SAS drives
- Network: Broadcom NetXtreme II5708 1Gbps
- vCenter Update Manager software: Version 5.0
- vCenter Server software: Version 5.0

NOTE: In addition to this hardware configuration with 16GB RAM and 4 CPUs, Update Manager and vCenter Server were deployed in virtual machines with 4 virtual CPUs and 4GB RAM. The virtual machine configuration achieved performance comparable to the hardware configuration.

ESX/ESXi System

- Host Computer: Dell PowerEdge 2900
- CPUs: Two 2.66GHz Intel Xeon 5355 quad-core processors
- RAM: 32GB
- Hard drives: Eight 73GB SAS drives
- Network: Broadcom NetXtreme II5708 1Gbps
- ESX/ESXi software: Versions 4.1 and 5.0

Virtual Machine Operating Systems

- Windows: Microsoft Windows XP SP2

Network Simulation Software

- WANem 1.2

Network Configurations

The network configurations used in the experiments are shown Figure 1 (basic network configuration) and Figure 2 (high-latency network configuration).

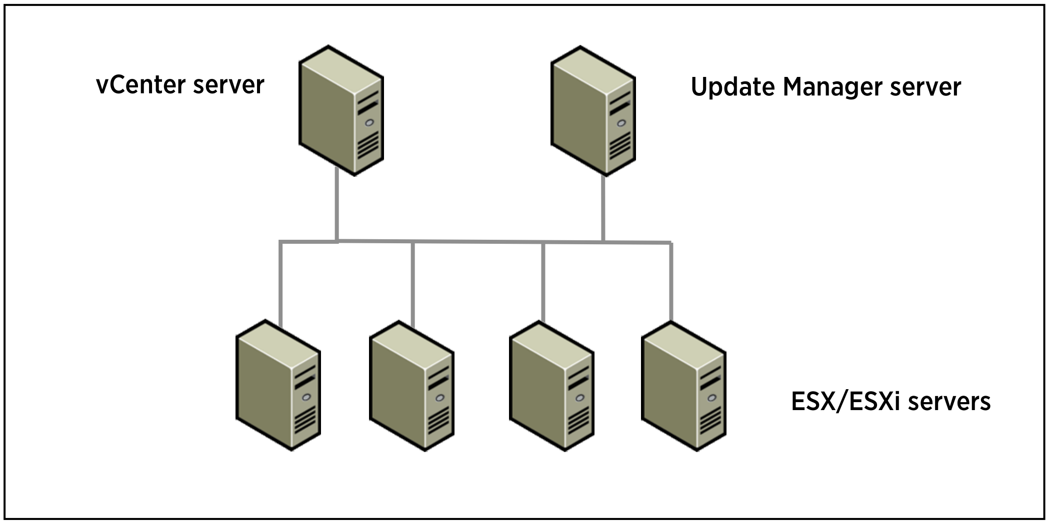


Figure 1. Basic Network Configuration

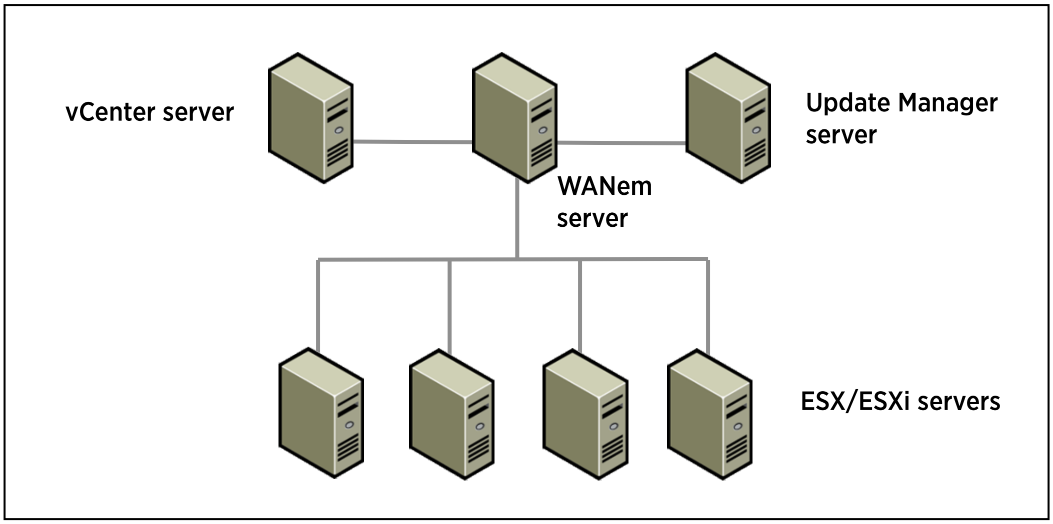


Figure 2. High Latency Network Configuration

Update Manager Server Host Deployment

There are three Update Manager server host deployment models as shown in Figure 3, where:

- Model 1 —vCenter Server and the Update Manager server share both a host and a database instance.
- Model 2 — Recommended for data centers with more than 300 virtual machines or 30 ESX/ESXi hosts. In this model, the vCenter server and the Update Manager server still share a host, but use separate database instances.
- Model 3 — Recommended for data centers with more than 1,000 virtual machines or 100 ESX/ESXi hosts. In this model, the vCenter server and the Update Manager server run on different hosts, each with its own database instance.

NOTE: If you are using Update Manager to update only hosts and not virtual machines, a system design where the Update Manager server shares the same host with vCenter Server is acceptable.

For both ESX/ESXi and virtual machine patching, the Update Manager server transfers patch files over the network. To avoid unnecessary disk I/O, it is ideal if the Update Manager server host can cache patch files, some of which are several hundred megabytes in size, completely within the system cache. To this end it is desirable for the Update Manager server host to have at least 2GB of RAM.

It is also best to place the patch store and Update Manager database on separate physical disks. This arrangement distributes the Update Manager I/O and dramatically improves performance.

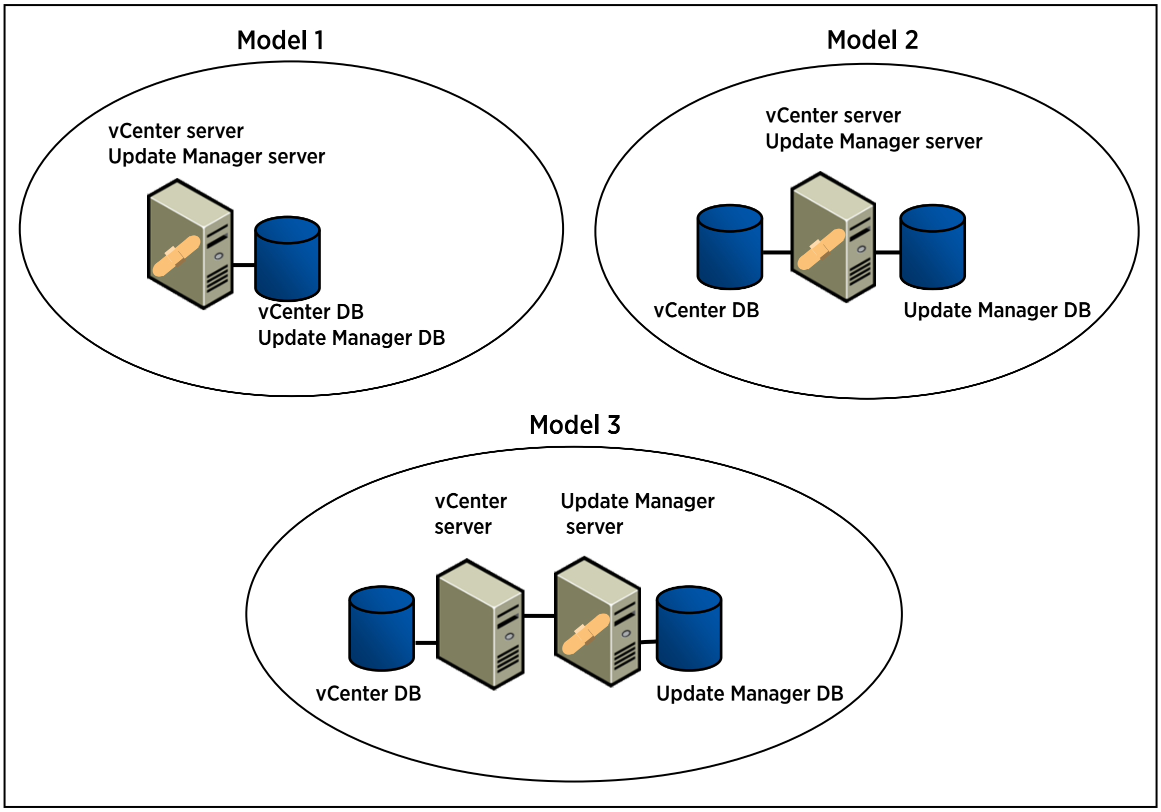


Figure 3. Host Deployment Models for Update Manager Server

Performance Tips

- Separate the Update Manager database from the vCenter database when there are 300+ virtual machines or 30+ hosts.
- Separate both the Update Manager server and the Update Manager database from the vCenter Server system and the vCenter Server database when there are 1000+ virtual machines or 100+ hosts.
- Make sure the Update Manager server host has at least 2GB of RAM to cache frequently used patch files in memory.
- Allocate separate physical disks for the Update Manager patch store and the Update Manager database.

Latency Overview

Update Manager operational latency is an important metric. IT administrators need to finish applying patches or updates within maintenance windows. Figure 4 and Figure 5 show latency results for an the following tasks:

- ESX/ESXi hosts scan
- VMware Tools scan
- Virtual machine hardware scan
- Virtual machine hardware upgrade, with powered on and powered off virtual machines
- Display of compliance view for a folder with 600 virtual machines
- ESX/ESXi host upgrade
- VMware Tools upgrade

For the virtual machine hardware upgrade latency result in Figure 4, the virtual machine was originally powered off in one case and powered on in the other. For VMware Tools upgrade, the virtual machine was originally powered on. For the display of a compliance view of a folder with 600 virtual machines, the default VMware Tools upgrade and virtual machine hardware upgrade patch baselines are attached to the folder and the latency is measured for fetching the compliance data for all attached baselines. The result in Figure 5 is for an upgrade from ESXi 4.1 to ESXi 5.0. All numbers are averaged over multiple runs. Note that these latency numbers are only references. Actual latency varies significantly with different deployment setups.

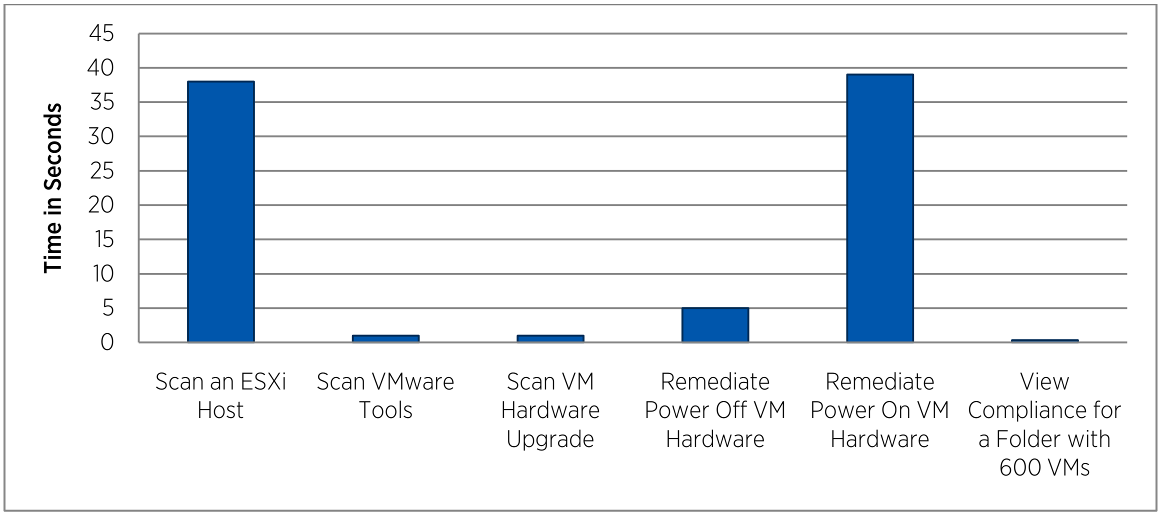


Figure 4. Update Manager Operation Latency Overview 1 (Latency in *Seconds* for Y Axis)

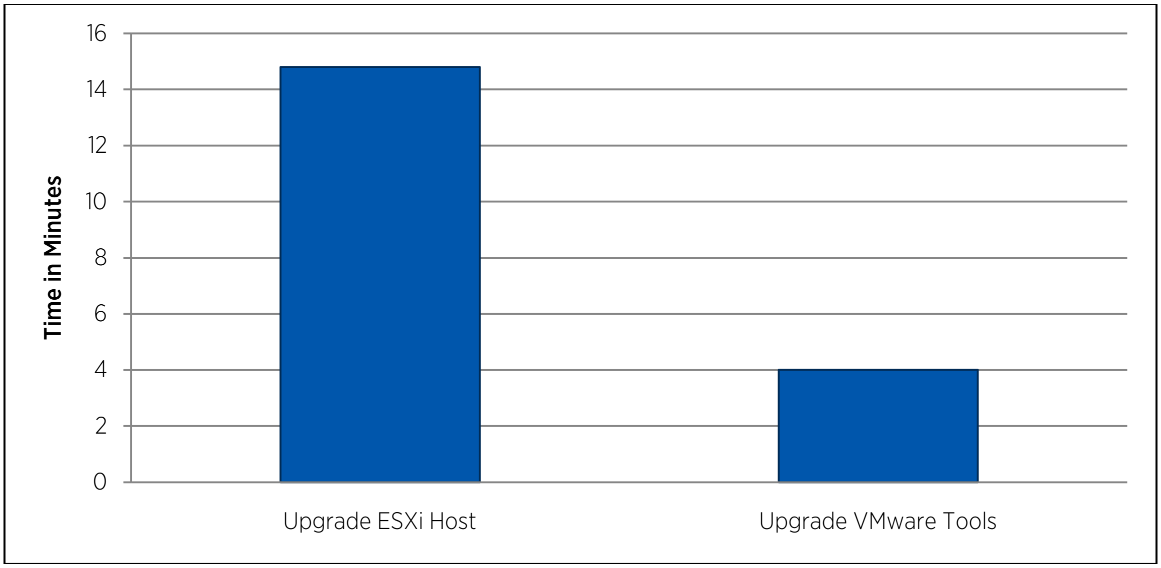


Figure 5. Update Manager Operation Latency Overview 2 (Latency in *Minutes* for Y Axis)

The results in Figure 4 and Figure 5 were measured on a low-latency local network setup. However, network latency plays an important role for most Update Manager operations. For some results obtained in a high-latency network, see the other sections in this paper, "Host Operations in WAN Environments" and "Network Throttling."

Performance Tips

- Upgrading virtual machine hardware is faster if the virtual machine is powered off. Otherwise, Update Manager powers off the virtual machine before upgrading the virtual hardware. This might increase the overall latency. *Note: Because VMware Tools must be up to date before virtual hardware is upgraded, Update Manager might need to upgrade VMware Tools before upgrading virtual hardware. In this case, the process is faster if the virtual machine is already powered on.*
- Upgrading VMware Tools is faster if the virtual machine is powered on. Otherwise, Update Manager powers on the virtual machine before the VMware Tools upgrade. This might increase the overall latency.

Resource Consumption and Job Throttling

Update Manager operations have different loads on the Update Manager server, vCenter Server, and ESX/ESXi hosts. Figure 6 divides Update Manager operations into low, medium, and high resource consumption levels.

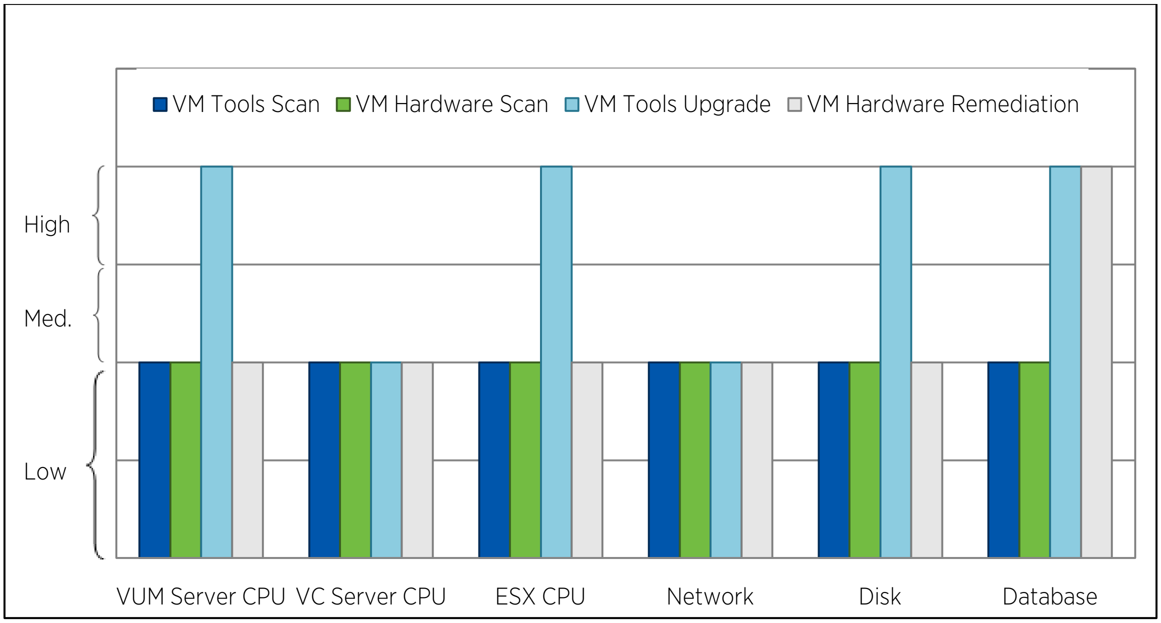


Figure 6. Resource Consumption for Update Manager

To alleviate resource pressure on Update Manager servers and ESX/ESXi hosts, Update Manager performs job throttling by limiting the maximum number of concurrent operations. Table 1 gives the default performance limits.

UPDATE MANAGER OPERATIONS	MAXIMUM TASKS PER ESX/ESXI HOST	MAXIMUM TASKS PER UPDATE MANAGER SERVER
ESX/ESXi host scan	1	72
ESX/ESXi host remediation	1	48
VMware Tools scan	72	72
VM hardware scan	72	72
VMware Tools upgrade	5	72
VM hardware upgrade	5	72
ESX/ESXi upgrade	1	48

Table 1. Job Throttling Default Limits for Each Update Manager Operation

Optimization for Cluster Remediation

Earlier version of Update Manager can only remediate a cluster of hosts in a sequential manner. In Update Manager 5.0, you can choose to remediate the cluster of hosts in sequence, in parallel, or in parallel with a given level of concurrency. You can choose the manner of cluster remediation in the VUM cluster remediation wizard. Remediating hosts in parallel can improve performance significantly by reducing the time required for cluster remediation. Update Manager remediates hosts in parallel without disrupting the cluster resource constraints set by DRS.

We use eight hosts in a cluster with DRS enabled for cluster remediation experiments. All eight hosts are the same model (Dell PowerEdge 1950) to ensure vMotion compatibility. The hosts are in a cluster with DRS enabled, so that virtual machines can be migrated with vMotion to other powered on hosts in the cluster if their original hosts need to enter maintenance mode or to be powered off.

Before the start of the cluster remediation, each host runs 10 powered on virtual machines. All virtual machines are consuming the same resource for each remediation experiment. However, for different runs of remediation experiments, virtual machines are set at a varying workload level (in our experiments, this is achieved by controlling the VM CPU usage).This way, we can observe the remediation performance with varying cluster resource usage.

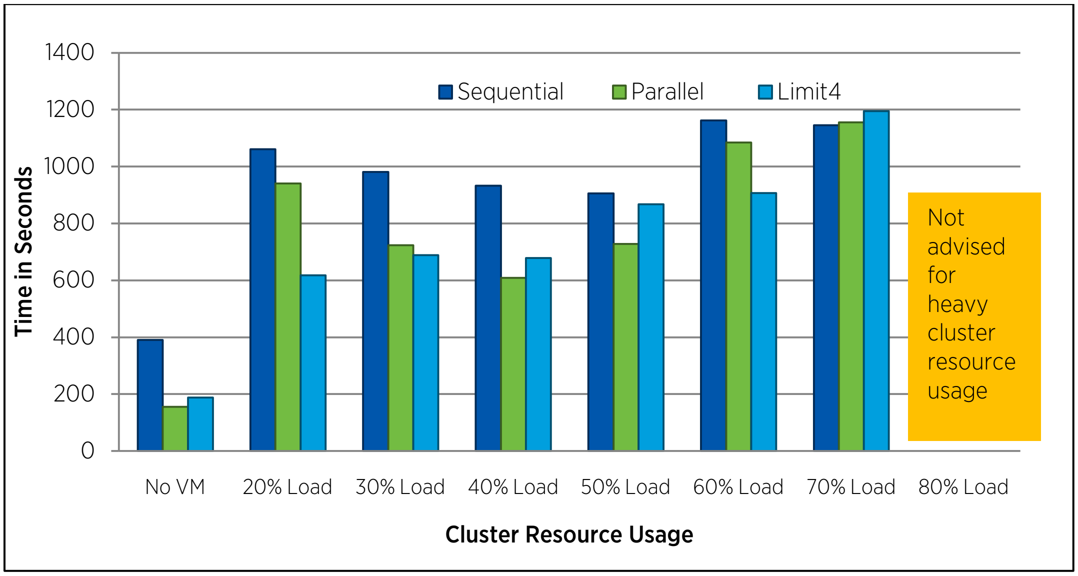


Figure 7. Time to Remediate a Cluster with 8 Hosts (Latency in Seconds for Y Axis)

Figure 7 shows the latency results on the experimental setup with varying workloads and different remediation methodologies, including in sequence, in parallel, or in parallel with a concurrency level of half the hosts' number in the cluster. It is observed that:

- For the case with no virtual machines, because no vMotion is required during remediation, the remediation time is less than for other cases, and parallel remediation improves performance.
- Parallel remediation improves performance because it allows concurrent remediation with virtual machines moved out of multiple hosts concurrently. For lower workloads, limiting the remediation concurrency level (that is, the maximum number of hosts that can be simultaneously updated) to half the number of hosts in the cluster can reduce vMotion intensity, often resulting in better overall host remediation performance.

NOTES:

- *In these experiments, small remediation patches are applied to each host, so most of the latency is caused by vMotion, which migrates virtual machines, in order to let the host enter maintenance mode. In reality, the remediation patch might be large and require a host reboot. In such a situation, more time is needed to remediate every host, therefore you might see more improvement by using parallel remediation.*
- *The time required for a host to enter maintenance mode is related to the number of virtual machines on the host that need to be migrated. In this experiment, before remediation there are 10 virtual machines with the same workload on each host. In reality the number of virtual machines with different workloads on each host might vary, and that is likely to produce different results.*

Performance Tips

- Limiting the remediation concurrency level (that is, the maximum number of hosts that can be simultaneously updated) to half the number of hosts in the cluster can reduce vMotion intensity, often resulting in better overall host remediation performance. You can set this option by using the Update Manager remediation wizard.
- When all hosts in a cluster have all virtual machines powered off and are ready to enter maintenance mode, concurrent host remediation will typically be faster than sequential host remediation.
- Cluster remediation is most likely to succeed when the cluster is utilized at no more than 80%. For heavily-used clusters, cluster remediation is best performed during off-peak periods, when utilization drops below 80%. If this is not possible, it is best to suspend or power off some virtual machines before the operation starts.

Host Operations in WAN Environments

If Update Manager is run over a low bandwidth, high latency, or lossy network, host operations might time out or fail for releases prior to Update Manager 4.0 U2. These operations include host patch and extension scanning, staging, and remediation, as well as host upgrade remediation tasks. Timeout might also occur under normal network conditions if the Update Manager operation takes more than 15 minutes to complete. Refer to KB 1021050, “Update Manager host tasks might fail in slow networks”.

For Update Manager 4.0 Update 2, and Update Manager 4.1 and later, no extra action is required and all host operations should complete without any errors. The time for host operations varies from several minutes (in LAN environments) to several hours (in WAN environments). You can estimate staging or remediation periods in various WAN environments as shown below. Table 2 lists typical WAN environments, such as dial-up, DSL, and satellite, with varying bandwidth, latency, and packet error rate.

NETWORK	BANDWIDTH	LATENCY	PACKET ERROR RATE
Dial-up1	64Kbps	250ms	0.05%
Dial-up2	256Kbps	250ms	0.05%
DSL	512Kbps	100ms	0.05%
Satellite	1.5Mbps	500ms	0.10%
T1	1.5Mbps	100ms	0.05%

Table 2. Bandwidth, Latency, and Error Rate for Common Wide-Area Networks

Figure 8 illustrates the maximum durations of host operations (upgrading to ESXi 5.0) in various network environments.

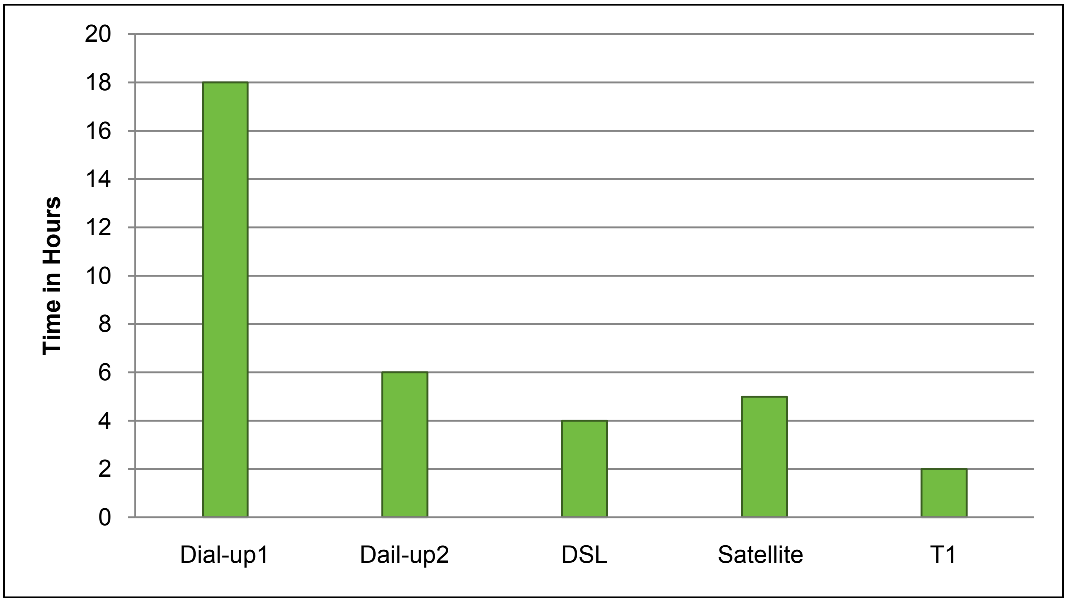


Figure 7. Maximum Host Operation Time

Performance Tips

- Upgrade to Update Manager 4.0 U2, 4.1, or 5.0 if you are working on a slow network.
- Host operations in a slow network take a long time. Refer to Figure 7 for the maximum time estimation. Do not interrupt ongoing operations.

Network Throttling for Host Staging and Remediation

In the Update Manager process, hosts download patches during remediation or while staging operations. To prevent patch downloads from using all available bandwidth in slow networks, you can configure ESXi 5.0 hosts to set bandwidth throttling. You can set a maximum value for downloading VIBs to ESXi 5.0 hosts from the Configuration tab of the vSphere Client, or by running an ESXCLI command. For more information, refer to *Installing and Administering VMware vSphere Update Manager*[1].

NOTE: Do not throttle the download bandwidth when you upgrade hosts. When you start an upgrade remediation, ESX/ESXi hosts are put into maintenance mode, and a limited download rate might cause hosts to remain in maintenance mode for an extended period of time.

The experimental results show that network throttling works well in various scenarios, including single or multiple network connections with multiple hosts.

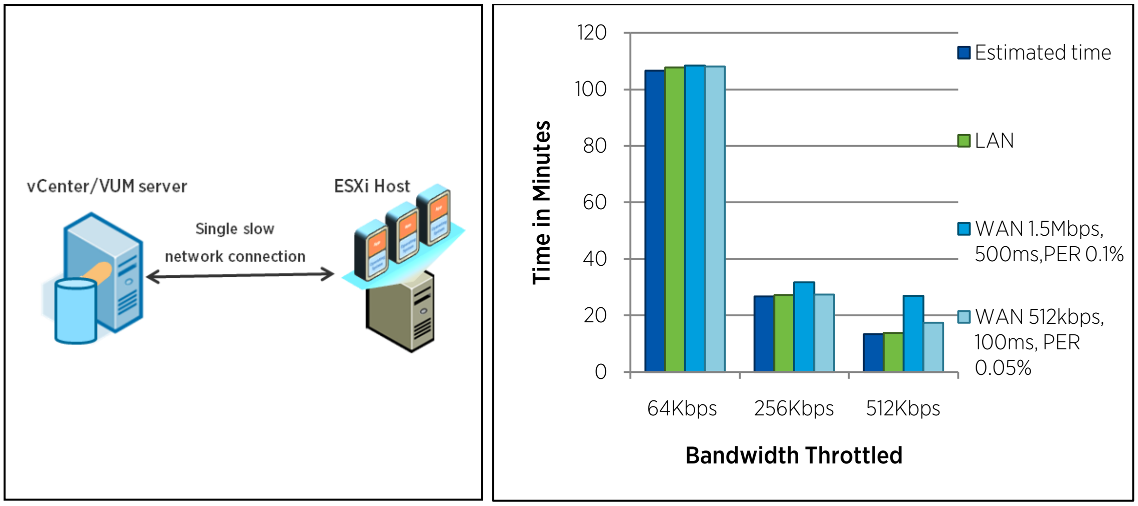


Figure 9. Single Connection, Single Host (Left: Experimental Setup; Right: Result)

Figure 9 illustrates network throttling performance in the simplest case with a single host and a single network connection. In this experiment, the ESXi host is downloading during staging a 50MB patch by using a single network connection that is configured as LAN and two slow network settings (WAN)—1.5Mbps/500ms, packet error rate (PER) 0.1% and 512Kbps, 100ms, PER 0.05%.

The X-axis shows the varying bandwidth throttled at 64Kbps, 256Kbps, and 512Kbps. The figure shows that the staging time is very close to the estimated time in either a LAN network or WAN environments. The results show that network throttling is performing well. However, when the network latency is high, the relative high bandwidth might not be achieved because of the effect of the bandwidth-delay product (BDP), which is the product of a data link's capacity in bits per second and its end-to-end delay in seconds. This effect appears in the medium-blue bar (1.5Mbsp, 500ms latency) for the 512Kbps throttled case, in which it takes more time than the other networks shown.

Network throttling also works as expected for the scenario when multiple hosts use a single network connection (Figure 10) and the scenario when multiple hosts use multiple network connections (Figure 11).

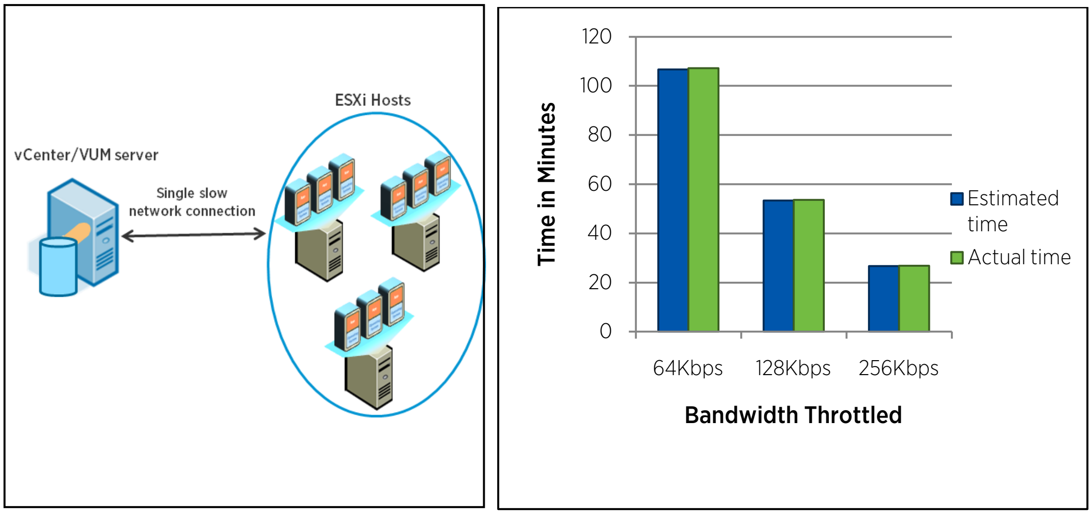
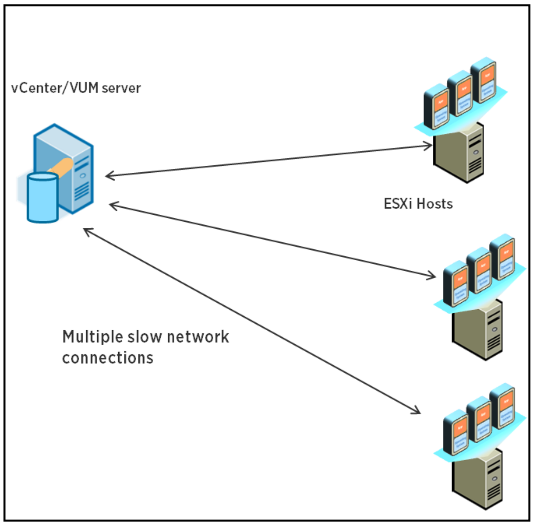


Figure 10. Single Connection, Multiple Hosts (Left: Experimental Setup, Right: Result)



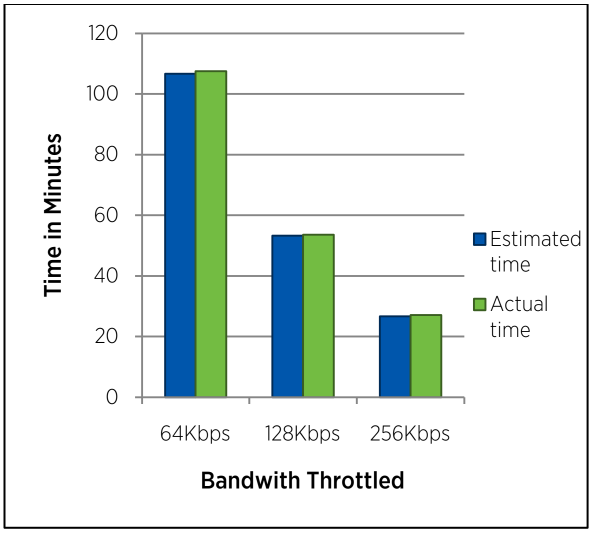


Figure 11. Multiple Connection, Multiple Hosts (Left: Experimental Setup, Right: Result)

For a single network connection, in order to ensure that network bandwidth is allocated as expected, the sum of the bandwidth allocated to multiple hosts on a single network link should not exceed the bandwidth of that link. Otherwise, the hosts will attempt to utilize bandwidth up to their allocations, resulting in bandwidth utilization that might not be proportional to the configured allocations, as Figure 12 shows.

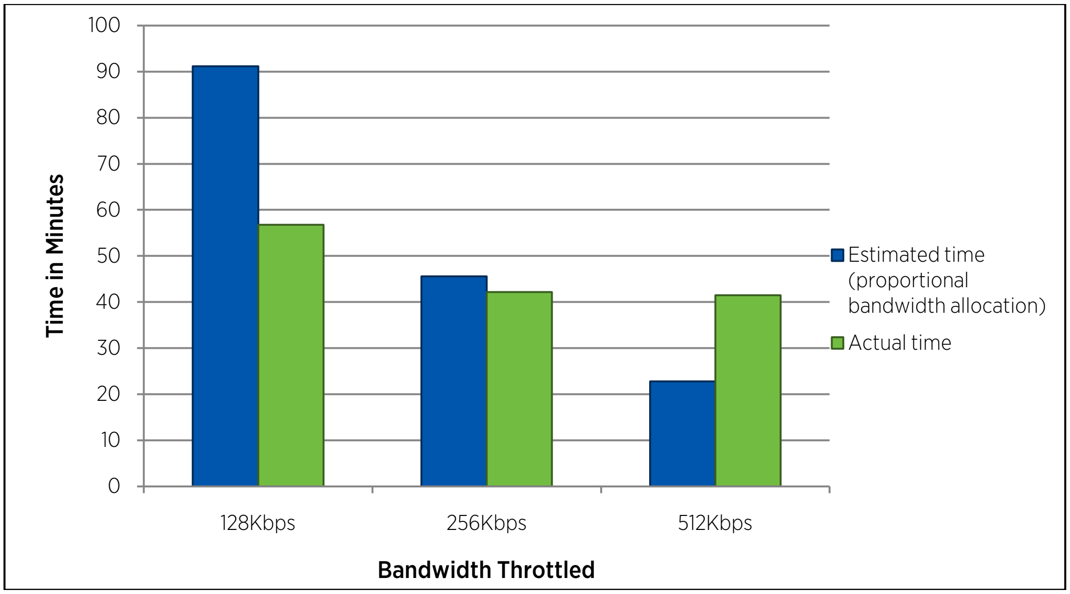


Figure 12. Single Connection, Multiple Hosts (Sum of the Bandwidth Allocated Exceeds the Link Bandwidth)

For example, in a single network with 512Kbps bandwidth, if the bandwidth throttling values are set to 128Kbps, 256Kbps, and 512Kbps for three hosts respectively, it is observed that the host with 128Kbps gets 128Kbps bandwidth, while the hosts with 256Kbps and 512Kbps only get about (512-128)/2 = 192Kbps bandwidth.

Performance Tips

- During remediation or staging operations, hosts download patches. On slow networks you can prevent network congestion by configuring hosts to use bandwidth throttling. By allocating comparatively more bandwidth to some hosts, those hosts can finish remediation or staging more quickly.
- To ensure that network bandwidth is allocated as expected, the sum of the bandwidth allocated to multiple hosts on a single network link should not exceed the bandwidth of that link. Otherwise, the hosts will attempt to utilize bandwidth up to their allocation, resulting in bandwidth utilization that might not be proportional to the configured allocations.
- Bandwidth throttling applies only to hosts that are downloading patches. If a host is not in the process of patch downloading, any bandwidth throttling configuration on that host will not affect the bandwidth available in the network link.

Conclusion

VMware vCenter Update Manager delivers the most full-featured and robust patch management product for vSphere 5.0. This white paper displays test data and recommends various performance tips to help your Update Manager deployments run as efficiently as possible.

References

1. *Installing and Administering VMware vSphere Update Manager*. VMware, Inc., 2011. http://pubs.vmware.com/vsphere-50/topic/com.vmware.powercli.vum_inst.doc_50/vumps_admgd_preface.html.
2. *VMware vSphere Update Manager Database Sizing Estimator*. VMware, Inc., 2011. http://pubs.vmware.com/vsphere-50/topic/com.vmware.ICbase/PDF/vsphere-update-manager-50-sizing-estimator.xls.

About the Authors

Xuwen Yu is a member of technical staff in VMware where he works on several projects including VMware vCloud Director, VMware vCenter Update Manager and host profile. Xuwen received his Ph.D. in Computer Science and Engineering from the University of Notre Dame in 2009.

Jianli Shen is a senior member of technical staff in VMware where she has been working on performance projects including VMware vCenter Converter, VMware View, VMware vCenter Capacity IQ, VMware vCenter Update Manager and host profile. Jianli received her M.S. in Computer Science from Georgia Institute of Technology.

John Liang is a senior manager at VMware. Since 2007, John has been working as a technical lead for performance projects on VMware products such as VMware vCloud Director, VMware vCenter Update Manager, VMware vCenter Converter, VMware vCenter Site Recovery Manager (SRM), and VMware vCenter Lab Manager. Prior to VMware, John was a principal software engineer at Openwave Systems Inc., where he specialized in large-scale directory server development and performance improvements. John received a M.S. in Computer Science from Stony Brook University.

Acknowledgements

We want to thank Rajit Kambo for his valuable guidance and advice. His willingness to motivate us contributed tremendously to this white paper. We would also like to thank Mustafa Jamil, Sirish Raghuram, Raj Shah, Kevin Wang, Jerome Purushotham, Roopak Parikh, Sachin Manpathak, Roussi Roussev, Cheen Qian, Ruopeng Ye, Muhammad Noman, Karthik Sreenivasa Murthy, Soam Vasani, Yufeng Zheng, Rajanikanth Jammalamadaka and the Update Manager team for their tremendous support. Without their help this white paper would not have been possible. Finally, we would like to thank Julie Brodeur for help editing and improving this white paper.

VMware, Inc. 3401 Hillview Avenue Palo Alto CA 94304 USA Tel 877-486-9273 Fax 650-427-5001 www.vmware.com