VERITAS

The Smart Use of Malware Scanning in NetBackup

Your organization's malware scanning tactics require an informed, proactive strategy. A strategy is needed because all IT environments have CPU, RAM, SAN, and LAN limitations, no matter how large or modern they may be. Just as administrators must be proactively strategic with backups, AIR replications, and SLP operations, they must be proactively strategic with malware scanning. Using NetBackup™ image scanning strategy is a secondary safety net to complement your primary on-system malware scanning. NetBackup malware scans are not designed or intended to replace your primary malware scanning. They are designed to help you restore clean, reliable data during disaster recoveries. To help you do so, this article covers malware scan host and sizing considerations for the NetBackup Malware Scanner, available to entitled customers from the Veritas download center. NetBackup Malware Scanner v2.1 and higher use parallelization to improve the scale of scan operations.

While administrators are easily tempted into thinking they need to scan every image for maximum malware protection, the key part of a proactive malware scanning strategy is to not immediately scan every image. Every scan operation ties up valuable resources and time. With data doubling (on average) every eighteen months or less, having enough time for scanning, regardless of other IT resources, becomes more of a daily factor. Therefore, only images that are highly likely to be restored or infected, or ones that will certainly need to be restored in a DR should be scanned at some point. That means malware scanning should be performed in alignment with your recovery point objectives (RPOs). This means your malware scanning strategy is modeled after your RPOs.

Efficient Malware Scanning

The timing of when an image is scanned is also extremely critical to an efficient scanning strategy. Once an image is created, any malware it contains cannot infect IT assets. This allows your scanning strategy the flexibility to scan where and when it makes the most sense before a restore of the data might be necessary. In smaller environments with enough IT resources to do so, scanning every new backup image between rounds of backups may make sense. Conversely, in larger environments, scanning only images selected by criticality to the organization may make more sense.

The likelihood and effects of malware in a backup image also determine if it should be scanned. The most likely to be restored images that also have a good chance of being infected should be scanned first. Conversely, the least likely to be restored or infected images should be scanned last or not at all. The structure of a data set also indicates how necessary it is to scan it. Unstructured data (also called unstructured information) does not conform to a specific data model and is not organized in a specific way, making it hard to sort through or read. Random files dumped on network attached storage (NAS) user shares are a great example of unstructured data. Unstructured data is the perfect place for malware to hide and execute using its host operating system. Structured data (also called quantitative data) is highly organized according to a standard or format, and therefore easily readable. Structured Query Language (SQL) database tables are a great example of structured data. Because of its rigid format, structured data is less likely to hold malware seeking to be anonymous and hidden, and in need of operating system execution.

Effective Malware Scanning

Here are some examples of effective, proactive scanning strategies for various organizations based on the timing and likelihood of ransomware in a backup image. You will need to similarly customize your scanning strategy for your environment resources and RPOs:

- A hotel chain uses lobby guest browsing systems that allow USB drive access. These systems are backed up locally for quick emergency restores, and to build a history of disk contents for later analysis of customer activities. The hotel chain's IT staff implement a strategy of scanning the first/primary copies of these images immediately, since an attack on them is almost inevitable.
- A data center administrator's strategy is to perform a final full backup of decommissioned VMs before the VMs are deleted from the hypervisor. Typically, such VMs are rarely restored and are kept purely to meet federal data retention requirements. They decide on a scan if needed for restore strategy for these images and apply their scan host resources to other images more likely to be restored.
- A web and service hosting data center has an RPO for their external customers that restores will begin within 30 minutes of data loss. To meet this RPO, they aggressively scan all full and incremental customer backups with as many scan hosts as possible so that every image is immediately restorable. They automate and control scanning with the NetBackup API immediately after backups are completed. However, for internal data assets, their RPO is to begin recovery within 12 hours, so their internal strategy is to scan only initial backup images of their most critical infrastructure. All other images are scanned before restoration as needed. Constant and aggressive infrastructure system malware scans are used to minimize the chance of those backup images being infected in the first place.
- A large military Tier 1 data center only hosts systems with no internet or USB access. All data center staff have thorough background checks and must pass security checkpoints to enter or leave the building. System backups are enormous, with millions of files and the best on-system malware scanning systems and procedures in place. Administrators decide that no matter what RPO they need to meet, they will only scan images right before restoration in case a new malware variant was identified before their malware scanners could detect it.

What Does the Average Customer Scan?

Performing regular scans of their most critical servers such as domain controllers, mail servers, and exceptionally large and mission-critical database servers is usually enough. Performing random spot check scans of other systems on a weekly or bi-weekly basis catches any threats that may have occurred. It is important to remember that ransomware/malware tries to spread to as many systems and as quickly as possible. The scenario of one non-critical system with unscanned backups being infected for a prolonged period while other systems remain uninfected is improbable. Malware is not designed to take out just a few systems—it is designed to move laterally within a compromised enterprise and rapidly cripple entire IT centers, forcing a mass shutdown or ransom payment.

Is it Ever Advisable to Scan Every Image All of the Time in Larger Environments?

From the examples and discussions above, many smaller or more powerful environments could scan every image—for now. Remember that data growth through day-to-day operations, let alone acquisitions and mergers, will change your scanning potential. Obviously, the smart move is to constantly reassess your scanning strategy. There are ways to augment your scanning capacity using DR assets to ensure your RPOs during a malware attack. Such scanning assets can be used after images are created and when unscanned images need to be restored. NetBackup's support of security information and event management (SIEM) platforms allows backup administrators to work with security teams to optimize your scanning strategy. You can read about NetBackup support of SIEM/SOAR/XDR systems here:

NetBackup 10.2 Expands Support of SIEM/SOAR/XDR

A terrific way to enhance scanning capacity while maintaining your RPO is with an isolated recovery environment (IRE) architecture using NetBackup. While scanning production backups is good, an IRE performs complete regular scans of data without impact to the production backup infrastructure. Using an IRE this way is documented in this solution paper:

FLEX Appliance IRE Solution

Augmenting scanning without an IRE requires bringing additional scan hosts online when needed. That is covered next.

The Smart Assignment and Use of Scan Hosts

Scan hosts are physical or virtual servers dedicated to scanning backup images for the best performance. That means they are not NetBackup primary or media servers, nor are they NetBackup clients backing up their own data. Instead, the scan host has a compatible malware scanner deployed that scans the stored backup images without needing to recover them onto an alternate system. Also, they must not be running any other applications besides the NetBackup components required for a scan host. Here is a short video showing how scan hosts are configured:

Veritas NetBackup Malware Detection Configuration

Here is a technical brief about how scan hosts work once they are in operation:

NetBackup Malware Detection Technical Brief

The video above demonstrates using build-your-own (BYO) physical or virtual systems as scan hosts. Both options have strengths and weaknesses. Physical systems can offer unshared and therefore higher network throughput than virtual systems in a hypervisor. This can be a huge advantage scanning network attached storage (NAS) system images, or exceptionally large backups comprising hundreds of thousands—or millions—of files. However, physical system RAM and CPU are not easily upgraded or moved between networks. Virtual system resources are changed quickly with a short outage/reboot, but must share their cluster node's networking with other virtual systems running on the same node. Despite these differences, physical and virtual systems can make high-performing scan hosts.

The smart move, therefore, is to plan your scan host creation and usage based on your organization's backup image data flow. For example, suppose you see large surges in quarter-end or year-end backups that must be scanned. In that case, you can temporarily use pre-built virtual scan hosts to augment that scanning load, then deprecate their resources and priority in the hypervisor. This leaves your normal NetBackup Appliance and physical BYO scan hosts to carry the scan load most of the time. It also minimizes costs and quickly ramps up scanning capacity only when needed. Therefore, like your overall scanning strategy discussed earlier, your RPO drives your scan host planning as well.

Malware Scanning Performance

Once you have considered all the points discussed above, you are ready to consider malware scanning throughput performance. Without the above planning strategies, even the best scanning throughput will not necessarily meet your scanning strategy and RPOs. Please remember that our test labs and data will differ from your NetBackup production environment. Carefully review all of the parameter and usage information below before estimating what scanning performance you will see in your environment.

First, there are some basic fixed scanning parameters to consider. The NetBackup release version being used determines what fixed parameters you have available:

- Instant Access (IA) is a critical part of malware scanning. The number of mounts per media server deduplication pool (MSDP) storage server is limited by NetBackup version and can vary across different supported form factors. Review your version of the NetBackup Deduplication Guide to verify how many IA mounts are possible in your configuration (such as 50 for NetBackup 10.2)
- The number of simultaneous scans per scan host was limited to three in NetBackup 10.1, and has been raised to 10 in release 10.2

Second, there are also variable/configurable parameters:

- Total number of scan hosts required for the scanning workload, and the recommended configuration for that workload
- Current MSDP storage server load stress on each MSDP server acting as an NFS/CIFS IA mount for a scan host
- Available network I/O throughput (network saturation) between the MSDP server(s) and the scan host(s) for the NFS/CIFS share(s) being used

Third, with 50 concurrent Instant Access mount points, you have a limit of 50 simultaneous backup image scans that can be started on any MSDP or WORM storage server. This assumes that all IA mounts are available for malware scanning. NetBackup 10.0 and 10.1 limit the number of parallel scans per scan host. To scan 50 images simultaneously with these versions, there need to be 17 scan hosts serving one storage server (3 scans x 17 hosts = 51 scans). This creates the possibility of over-subscribing the IA mounts on the MSDP server by one scan. More than 50 IA mount points can be created, but are not supported. Any additional load on the MSDP server during the scans results in a loss of overall MSDP performance. If there is little or no IA latency, no other MSDP load, and no network saturation issues, 17 scan hosts per MSDP pool having at least (but hopefully) more than 8–16 CPUs should handle the typical scanning workload.

In NetBackup 10.2 and above, the number of parallel scans is configurable. A maximum of 10 simultaneous scans can be started on each scan host. Like previous versions, the number of concurrent IA mount points per MSDP server is limited to 50 from an MSDP server or from a WORM storage server. This reduces the number of scan hosts required to scan at the maximum capacity of an MSDP server. Only five scan servers are required to use all 50 mounts with no chance of over-subscription (10 scans x 5 hosts = 50 scans). Flex appliances limit IA mounts to 50 from each node. This means that multiple Flex deduplication pools are ideally on different nodes, or the 50 mount points need to be manually managed across multiple WORM/MSDP instances on the same node.

Fourth, there are also IA tuning parameters available. IA performance is a key part of scanning performance. The IA parameters available are documented in the Security and Encryption Guide of each NetBackup release. The 10.2 version is here:

NetBackup Security and Encryption Guide Instant Access Tuning Parameters for Malware Scanning

Releases 10.0 and 10.1 have fewer tuning options. Be sure to review the Security and Encryption Guide for your NetBackup release so that you know what tuning options are available to you.

Creating a Test Plan and Rollout for Malware Scanning

It is critical you create and execute a limited malware scanning test program before implementing this feature widely in production. You do not want to over-commit (waste) or under-commit scanning resources. The resources you commit will vary with the file counts and sizes of the images to be scanned. Testing will also identify any unexpected networking or server issues you may have. It will also benchmark what typical scanning performance is in your environment, by which you can plan the requirements of your full production scanning implementation.

Be sure you are using the latest recommended malware scanner software in your testing. That is version 2.2 as of this posting. Be sure to review your NetBackup documentation and the Services and Operations Readiness Tools (SORT) site for more information on which scanner software may be available and how to upgrade yours to the latest version, if needed.

Most importantly, make sure you understand the requirements for a scan host. That information is available at these two links:

NetBackup Security and Encryption Guide, Prerequisites for a Scan Host

NetBackup Security and Encryption Guide: UNIX, Windows, and Linux

Here are the scan host recommendations as of release 10.2:

- A dedicated physical or virtual system with eight CPUs and 32 GB of RAM. Scan hosts with up to 16 CPUs and 64 GB of RAM may be required in very demanding scanning situations where images have hundreds of millions of files
- NetBackup BYO Deduplication Media Server or the WORM storage server systems on the Flex appliance platform should have compute and throughput capabilities comparable to NetBackup Appliance or Flex Appliances, or better

- As stated earlier, in release 10.2 the number of allowed scan host jobs default to 3, are limited to a minimum of 1, and a maximum of 10 per scan host. These settings are bounded by NetBackup software and cannot be outside that range. In both testing and production, no more than 50 scans should be running simultaneously against one storage server. We recommend using no more than 20 concurrent malware scans on systems that are also utilized actively for backups or duplication jobs. This can be achieved using a combination of the number of scan hosts times the maximum concurrent scans defined per scan host. Staying at 20 scans or below in daily production reserves IA mounts for IA restores, or large-scale scanning during disaster recovery operations. Mounts used for malware scanning should fully dedicated to just scanning. No scan mounts should be used for anything else (like restores) during scanning testing or production

You will need more scan hosts for backup images containing large numbers of files than you will for large backup images with few files. Backups with exceptionally large file counts are split into smaller batches of 500,000 files each and scanned in parallel. Each batch creates a separate Instant Access mount point and counts toward the IA limit per storage server. To illustrate, to scan a backup with 20 million files, 40 batches will be created and if the user wants to limit IAs to 20 on the storage server, two scan hosts (with 10 parallel scans each) would suffice. Additional scan host(s) can also be provisioned within the scan pool, but marked inactive. These are helpful if we need to swap the active scan host during a maintenance window or outage.

Document your test results comparing backup image size(s) and file count(s) processed, with scanning throughput times. Your results will indicate how many scan hosts are required for the workload(s) you will be scanning. Have a replacement plan for scan hosts if one or more fail. You can also provision additional scan hosts in a pool and toggle activation status; this is helpful if you want to perform any maintenance on a scan host and need another one handy. In production, start with a limited scanning scenario such as the one(s) you did in testing. If that works well, begin slowly scaling up scanning, checking for unexpected issues. Do not try to scan all images immediately in any NetBackup domain, no matter how small the environment may be. As you scale up production scanning, document your results to establish what a normal/expected level of scanning performance should be. Most importantly, add malware scanning performance to your regular reporting for team review. Your malware scanning requirements and performance will certainly vary as your data changes in size and content over time.

Summary

Used wisely, NetBackup malware scanning is a terrific way to augment your on-system production malware scanning. It is a critical part of ensuring DR restores are clean and will not reinfect systems. It also ensures your most critical backups are clean in storage and ready for immediate restoration.

Scanning performance is dependent on these factors:

- Size of image
- Number of files in image
- Number and compute resources of MSDP storage server and scan hosts
- Disk, LAN, NFS, IA performance between scan hosts and NetBackup storage units

NetBackup scanning strategies are implemented based on RPOs. With the proper resources and implementation, NetBackup malware scanning is an invaluable tool to ensure you meet your RPOs without worrying about reinfecting your systems.

VERITAS

2625 Augustine Drive
Santa Clara, CA 95054
+1 (866) 837 4827
veritas.com

For global contact information visit:
veritas.com/company/contact

V1911 05/23