Insight Frequently Asked Questions

version 2.0 (8/24/2011)

Insight Overview

1. What is a reputation system and how does it work?

Insight, our reputation system, leverages anonymous telemetry data from over 150 million contributing users worldwide to automatically assign a reputation rating to every application file (e.g., EXE file) on the internet. Our system harnesses the "wisdom of the crowds" to automatically assign safety ratings to files. Our system doesn't just track bad files, and it's not just a bunch of fingerprints that our analysts post to servers. Instead, it contains real-time safety ratings on every known executable file, GOOD or BAD, across the world. Symantec's security products now leverage Symantec's reputation data in order to block new attacks, to reduce false positives, and to improve the performance of our products.

2. What types of files does the Insight system cover?

Our system assigns reputation ratings to all Windows executable files including traditional EXE files, driver files (including printer drivers), screen saver files, DLLs, OCXs, MSI Installer files, etc. and has reputation data on literally billions of unique executable files, both good and bad. We do not currently use reputation to convict other file types such as macros within Office documents or PDF files. For these types of files, we rely on our other security technologies like AntiVirus, SONAR and Intrusion Prevention which are well suited for detecting and convicting threats found in complex document formats.

3. How is Insight leveraged in the upcoming SEP 12.1 product?

Reputation is used in four different ways:

A. Each time the end-user downloads a software program, SEP will check the reputation of the file before the user is allowed to run the program. Downloads with a bad reputation are blocked automatically. SEP can block downloads based on their reputation, their prevalence or their discovery date. The administrator may specify different blocking thresholds in SEP policy on a per-user or per-group basis. Thresholds can be based on the reputation of the file, its age, or its prevalence. For instance, the administrator could specify that for the Finance client group they want to block all low reputation software as well as all software that is less than 1 day old or which has less than 100 users worldwide. However, for the Helpdesk client group, they want to allow medium-reputation software with at least 10 users. Given that most malware is generated on-the-fly for each victim, such powerful policies can help corporations block the vast majority of new malware coming from the web, since these single instance threats have low prevalence and short life spans.
B. Each time the user performs a traditional virus scan of their computer (e.g., a weekly scan) or our AutoProtect real-time component scans a file, it will first determine if the file appears to be suspicious. If the file does not appear to be suspicious, no reputation lookup is

performed. On the other hand, if the file does appear to be suspicious to our heuristics/fingerprint scanner, SEP looks up its reputation. This amounts to a very small subset of all files scanned. Any file which has suspicious characteristics and also a low reputation will be removed from the system. On the other hand, a file with suspicious characteristics but a good reputation is allowed to run unfettered, reducing false positives.

C. Each time the user runs a program, and the program exhibits suspicious behaviors, our SONAR behavior-blocking technology will look up the program's reputation before blocking these behaviors. If a program exhibits suspicious behaviors and also has a low reputation, it will be blocked. Programs that exhibit suspicious behaviors but have a good reputation are allowed to run, preventing false positives.

D. While the endpoint is idle, SEP 12.1's new ScanLess feature inventories all actively running applications and looks up their reputation (there are typically under a hundred active applications running on an average system, so this represents very few queries). SEP than saves these reputation results in its secure database. These files are never scanned again, even after fingerprint updates for our fingerprint scanner, improving performance and reducing false positives. Should any one of these files change (even by a single bit), then they are automatically rescanned by SEP.

4. <u>What happens if your reputation has an incorrect rating on a file?</u>

Symantec has the ability to manually override every reputation rating. If we receive a dispute from a software vendor or customer, we can usually adjust the file's reputation within a 24-48 hour window.

5. <u>Can malware authors poison the reputation database by, for example, getting lots of people (or bots) to download a posted file, therefore increasing its positive reputation?</u>

All reputation-based solutions have to be aware of attempts to undermine the system. Just like Google's PageRank system, we have to take this sort of attack very seriously. The reputation system takes this into account and has a number of "anti-thwarting" features to prevent such a scenario. For example, we will not honor submissions that do not come from a customer with a valid license. For a malicious user to poison the system, it would be necessary for them to submit the same file from a large number of different systems all with valid licenses. Additionally, those systems would have to have a good reputation for the submissions to have enough weight to affect the file's reputation score.

6. <u>Do we outright trust files with digital signatures?</u>

No. While each of our protection technologies checks for digital signatures and uses this data in their decision-making process, the simple presence of a digital signature will not, on its own, cause a file to be trusted.

Some technologies (e.g. SONAR) use a digital signature as a positive heuristic for false positive (FP) mitigation in certain workflows. Also, we attempt to ensure that our signature based technologies do not FP on signed binaries which are critical to system operation. However, a file that is signed by an untrusted third party can still be detected and remediated by our technologies, especially by our traditional signature based detection engines.

7. Since hashes are used for fingerprinting, is there a risk of hash collision in our database?

Our system uses 256-bit SHA2 hashes to identify files. To date, the SHA2 algorithm has never been broken or successfully attacked by cryptographers and it is considered a cryptographically secure algorithm (see http://en.wikipedia.org/wiki/SHA-2).

8. Does Insight do “whitelisting”?

Since Insight has reputation information about both good and bad files, we are able to “whitelist” good files that may appear suspicious. This means, for example, that if SONAR sees a process exhibiting suspicious behavior, but the executable is known to be a trusted app in our reputation database, SONAR will not convict the file.

9. How can a customer get a file or set of files added to our cloud reputation database?

The SEP 12.1 product ships with reputation information for the most popular applications known to Symantec; this includes virtually all Windows system files and top-tier applications. No internet lookup will be performed on these files and they will be automatically trusted (and never scanned), speeding up performance.

In addition, the customer can also add exclusions for reputation convictions within their organization in one of three ways:

A. The customer can use the SEPM console to locally whitelist files (e.g., internally developed applications) that are known to be clean. Our product will never issue reputation lookups for these files once they have been whitelisted. These files will only be whitelisted within the customer environment and this data is not submitted to Symantec or any other users.
B. The customer can specify that files downloaded from their intranet should not be convicted based on reputation alone.
C. The customer can contact Symantec to request that files be added to the global reputation database. Symantec now has a Reputation Submission Tool that can be provided to your organization which can be used to submit hashes of trusted files, as well the software files themselves (if you’re willing to submit these) so they can be added to our whitelist database. (See https://submit.symantec.com/whitelist/bcs.cgi for more information.)

10. What differentiates the way SEP does reputation vs. the competition?

Many of our competitors say they have file reputation systems. These “reputation” systems are typically offering malware “signatures” in the cloud. In other words, they function by comparing files against a list of known malware – a list that is maintained on a server instead of on the computer to be scanned. This approach allows for faster reaction to new malware and, in some cases, for faster scans.

However, cloud based blacklists still require the security vendor to collect a new virus sample, analyze it and fingerprint it before they can offer protection via their cloud servers. They are unable to protect against entirely new threats that they have not received, analyzed and fingerprinted in their lab. In other words, their blacklists in the cloud react faster – but they are still reactive, can't address new threats, and are focused on identifying malicious applications.

In contrast, our reputation system has accurate safety ratings on every single executable file known to any of our 150 million customers which currently has reputation data on literally billions of unique executable files, both good and bad. Our system doesn't just have fingerprints for files that our researchers have manually analyzed in our labs; it has user wisdom-derived ratings for all files. These ratings are calculated automatically as we receive submissions from our customers. By providing reputation for both good and bad files, we are not only able to detect malicious files which we have never seen before, but we are able to confidently identify good files as clean.

Insight Performance

11. <u>We say Insight can reduce AV/AS scanning by up to 80%, but it only reduces scanning for Portable Executable files and other potentially malicious file types. So if a machine is heavily loaded with files of other types, wouldn't they all have to be scanned?</u>

This is correct. Insight primarily eliminates scans on high-reputation executable files, so it is not likely to have a performance impact on an on-demand scan of other types of files (e.g., DOC or MP3 files). That said, Insight does help to reduce real-time scans by 80-90%, meaning that overall, every-day computer performance will be substantially improved. This is because while only 24% to 38% of files on the average computer are portable executable files, these are the files that are most frequently scanned by our real-time scans.

12. <u>When new AV/AS definitions arrive, wouldn't previously scanned (and subsequently trusted) files have to be scanned again?</u>

Trust is gained by having a high reputation, not by our antivirus scanner having scanned a file before. Thus a Symantec-trusted file will not be scanned even when new definitions arrive.

Insight Network Usage

13. <u>What type of network communications are required for your reputation system to work?</u>

Our reputation system uses two types of network communications, submissions and lookups. All data is transmitted over HTTPS and is cryptographically encrypted.

A. **Submissions** – Our SEP 12.1 product submits information about the software applications running on each system to Symantec. These submissions occur while the machine is idle (e.g., the user is out to lunch) and all information is anonymized to prevent private information from being sent. Typical information sent includes the SHA2 hash, pathname, download URL (for internet-downloaded files) and creation date of each executable file running on the system. The actual files themselves are not sent to our reputation service. In addition, our product sends information on previous infections encountered on each PC (e.g., this machine had 5 infections last week). Symantec uses this telemetry data as input to its reputation algorithms – the more users that submit data, the better our reputation algorithms work. It is a community effort. The submission of one benefits many. It is not strictly required for your corporation to submit data to our reputation service for our reputation system to work – you can disable submissions without any loss of protection; however, submitting data helps us to improve our accuracy on the files used in your organization.

B. **Lookups** – As mentioned in the sections above, during typical computer operation, any time our product encounters a potentially dangerous or suspicious file (such as a new download or a program behaving erratically), our product looks up the file in question in our cloud-based reputation service. These lookups also contain information such as the SHA2 hash of the file in question, the URL where the file was downloaded from (if any) and the pathname of the file. We use this data to lookup the file in our reputation database in real-time and then send the latest reputation rating back to the client. Our client product then uses the result to decide whether or not to block the file in question. Unlike submissions, reputation lookups cannot be disabled without degrading the level of protection offered by all security technologies in the product.

See questions 15 and 16 below for specific bandwidth usage information.

14. <u>How frequently will your product perform reputation lookups?</u>

This varies based on use case:

A. **Download Advisor:** SEP looks up the reputation of all downloaded software files. Given that the average user downloads less than 1-2 new apps per week, this will result in perhaps 1-2 queries per week.

B. **Traditional virus scan:** Our fingerprint scanner looks up the reputation of programs that appear to have suspicious instructions or a fingerprint of a known threat. Reputation queries do not occur for good, clean executable files. This amounts to roughly 0.5% of all programs scanned during a typical scan, or roughly 3-5 reputation lookups per full-drive scan.

C. **SONAR:** Our SONAR system looks up the reputation of only those programs that exhibit sufficiently suspicious behaviors, not for executable files that are considered clean. This amounts to less than 1% of all programs running on a typical system, with perhaps 1-2 reputation lookups per week.

D. **ScanLess:** Our ScanLess feature looks up all actively running apps to determine their reputation; it only rechecks applications with a low reputation, avoiding frequent redundant reputation checks on high-reputation items. You can expect that right after the first installation, ScanLess will look up approximately 100-200 files on a typical system during system idle periods. After these files have been looked up once, the system will re-check the reputation of roughly 1-2 new files every week (or as often as new applications are installed).

Altogether, our data shows that after a client has done its initial population (via ScanLess) the average client performs a total of 1-2 lookups per day.

15. What is the aggregate bandwidth used for reputation lookups (and submissions) right after install?

After a client has installed the product, we will have our first chance to review the files that are installed and running on the machine. We will typically send between 300 kilobytes and 1 megabyte of data during the first few weeks. In some extreme cases, we may send as much as 3 megabytes of data depending on the amount of software installed and actively in use on the machine. Submissions will take place when the machine is idle, but queries will be scattered across the period of time as our detection technologies analyze files for the first time.

16. What is the average daily/weekly bandwidth used for reputation lookups (submissions) during steady-state operation of my typical PC (a few weeks after installation)?

Once the initial round of submissions and queries have completed, the amount of data will be fairly minimal. New submissions and queries will only be sent when new files arrive on the machine, either through downloads or installation of other software. You can expect approximately 4 kilobytes of data for each new executable file placed on the machine to be sent to Symantec. As mentioned above, our data shows that the average machine performs 1-2 reputation lookups per day. This equates to approximately 4-8 kilobytes of transmitted data per day.

17. When I first deploy SEP to my machines and a scheduled scan occurs for the first time, will this cause a mass of Internet connections?

In practice, the amount of bandwidth used during scans should be negligible. Only those files that appear suspicious to our scanner have their reputation looked up. In practice, this amounts to roughly 0.5% of all scanned executable files, or about 3-5 files on a typical Windows XP or Windows 7 system. In addition, assuming multiple files need to be looked up, our system doesn't look them up one-by-one. Instead, our engines batch multiple lookups together to reduce the number of internet connections generated during a scan.

18. If a customer disables reputation lookups to Symantec's cloud, approximately what % of protection is lost?

If you disable reputation lookups in your environment, SEP 12.1's level of protection will still exceed that offered by the previous versions, due to a number of improvements to our protection stack. We believe that even without reputation, SEP 12.1 offers best-of-breed protection when compared with any of our major competitors. That said, disabling reputation lookups will definitely have a negative impact and reduce protection. The following is an assessment of the impact on every layer of our security stack:

A. For Intrusion Detection, only a small fraction of protection will be lost.

B. Our new Download Advisor component can't function without performing reputation lookups. Therefore, disabling reputation lookups effectively disables Download-based reputation protection.

C. Some of Symantec's heuristic detections (non-signature-based, generic file scanning technologies) have reduced effectiveness if reputation lookups are disabled. Why is this? A subset of our newest heuristics was designed to be extremely sensitive. As such, occasionally they will incorrectly flag legitimate software files as being suspicious. To ensure an ultra-low false positive rate for these sensitive heuristics, our product will not block these suspect files unless it has also looked up the file's reputation and verified that the reputation is poor. If reputation lookups have been disabled then our product will not be able to determine this and cannot safely block these suspicious files, so the scanner ignores these detections.

D. For our SONAR behavior blocking technology, turning off reputation lookups will significantly impact the detection rate (for the same reasons cited in the paragraph above). Additionally, if submissions are also disabled in addition to reputation lookups, then there will be loss of visibility from Symantec into threats targeting your organization, which will mean that we will not be able to respond as quickly to tune the protection for your organization.

19. <u>Each SEP client can submit info about new files, is there an overhead on the LAN?</u>

There should be little impact on the typical LAN due to either reputation submissions or lookups. (See question 15 above on bandwidth consumption).

20. <u>What protocol does Insight use and is it secure?</u>

Insight uses the HTTPS protocol and sends all data over secure sockets.

21. <u>Can the Insight functionality be provided to endpoints without the endpoints sending lookup requests directly back to Symantec, perhaps through the use of a server containing replicated reputation data within my organization?</u>

While some reputation information is cached on each client, reputation lookups for newly downloaded files require a connection to Symantec. We are researching the ability to allow an organization to locally cache reputation lookups, but that is not available in SEP 12.1.

22. <u>Is there any specific firewall configuration required to talk to the Symantec reputation system?</u>

Most firewalls already allow Insight queries. Queries currently use the HTTPS protocol over the standard port 443 to a Symantec controlled server.

23. <u>Are reputation submissions and lookups proxy aware?</u>

Yes, reputation submissions and lookups use the same proxy settings used for other technologies within SEP.

Insight Data Content

24. Is Personally Identifiable Information (PII) data sent during either lookups or submissions?

No, PII data is not included in reputation lookups or submissions. Currently, the only textual data sent in lookups and submissions are the pathname (e.g., “c:\program files\acrobat reader\reader.exe”) of each file and the source URL of downloaded executable files (e.g. http://virus.com/foo.exe”). Other data in reputation submissions and lookups include

- The SHA2 hash of the file,
- The local creation date of the executable file,
- The hash of the “parent” file, or the process that created the file on the local disk (e.g. iexplore.exe), and
- Information about the digital signature of the executable file, if applicable.

25. What do you do with all submitted data?

All submitted data is stored in secured Symantec data warehouse on-premise in Symantec data centers. Access to submitted data is strictly limited.

26. Do you track where reputation submissions and lookups come from?

No, all client IP address information is discarded at submission time and is not stored along with our reputation data. Our goal is to ensure that all participating users are entirely anonymous.

27. Can my organization see all the data sent to Symantec?

At this time that is not available. Currently the data is sent directly to Symantec in encrypted form.

28. Where exactly is the reputation data stored and for how long is it retained?

On the client, reputation data is stored in a securely encrypted local database on the client machine. Each entry in the database has a unique time-to-live value which determines the point at which the data would need to be refreshed. The data will only be refreshed on demand, though. So a binary which has an expired time to live will only be refreshed the next time a request is made for information on that file.

On our servers, customer-submitted reputation data is stored in a secure Symantec data warehouse.