The National Institute of Standards and Technology (NIST) National Software Reference Library (NSRL) has created curated releases of the Reference Data Set (RDS) consisting of hashes of Kaspersky products. This is in response to the DHS directive on Kaspersky applications: https://www.dhs.gov/news/2017/09/13/dhs-statement-issuance-binding-operational-directive-17-01

The data set includes the products listed in the directive plus additional Kaspersky products from the NSRL collection. System administrators and IT personnel tasked with implementing the directive may use this RDS release with third party administrative tools to discover the presence of potential Kaspersky installations.

The current release, comprised of information derived from Kaspersky installation media, may be found at:
https://s3.amazonaws.com/rds.nsrl.nist.gov/kaspersky/current/RDS_Kaspersky.zip

The RDS only contains hashes from installation media. The NSRL is supplementing this with hashes from an alternative methodology to derive additional hashes. The NSRL has applied their Diskprint process, with some modifications, to these Kaspersky products in order to observe file system artifacts generated through application use. A complete description of the Diskprint hashset follows these links.

The Diskprint project is described at: https://www.nist.gov/itl/ssd/cs/diskprints

The Kaspersky data in several formats---including the NSRL's RDS format, the Software ID (SWID) tag format, and the Digital Forensics XML (DFXML) format---and supporting data is at:
https://s3.amazonaws.com/rds.nsrl.nist.gov/kaspersky/current/Diskprint_Kaspersky.zip

NIST may provide periodic updates to the data sets described in this document.

The principal goal of producing Kaspersky data with Diskprinting is identifying only consistently-appearing files that are associated with the specified Kaspersky products. The file signatures provided here should not match on a system where no Kaspersky product was used; toward that goal, the data have been filtered with some probability of false negatives (i.e. missing files truly associated with Kaspersky software), but in exchange there should be no false-positive associations (e.g. declaring an operating system file as belonging to a Kaspersky product). Should this prove to not be true in practice, feedback is welcome at nsrl@nist.gov.

To start making the data, the NSRL deployed the Diskprint process, with one deviation. The Diskprint process usually follows a minimal-interaction approach to creating lowest-common-denominator software footprints: Software is installed, opened, and immediately closed on snapshotted virtual machines, in order to generate artifacts one will find if software is ever run on a system. For the Kaspersky prints, a more free-form software interaction pattern was devised and followed for each product, generating artifacts from somewhat typical, instead of minimal, usage. Each product was diskprinted once to exercise its capabilities (especially simple virus scanning and signature store updates), and then diskprinted twice more following the steps taken the first time as closely as possible, usually within the same business day. This triplicate printing created file system artifacts that could be observed to appear consistently or not.

To identify files that appear consistently, only files that appeared when exercising the software in all three out of three prints were considered. If a file's content appeared three out of three times, its hash

is reported without qualification; if it appeared at least two out of three times, the hash is reported, but flagged (in formats that support flagging) to indicate the hash might not be expected to appear in an administrator's scan of future systems.

To exclude files believed to not be associated with only the Kaspersky products, two processes were followed. First, the full paths of file names were manually inspected, and path prefixes were identified as indicators of files that were not present due to just the Kaspersky products (such as Windows log file directories). These path prefixes are available in the "Supplemental" directory of the Diskprint zip. Second, any file with a hash found in the NSRL's RDS and not associated with a Kaspersky product was excluded. The RDS used was the Modern and Legacy distributions, version 2.58; the Kaspersky products are listed in the first RDS link above.

The filtered file signatures are then converted into several forms, linked to above. The variety of forms is due to some formats supporting different metadata facets that administrators may find useful. The NSRL RDS format does not support file path information, but has long been supported by various forensic tools. The SWID format has a principal use case of supporting software inventorying, and supports some organizational metadata and file path linking to common Windows environment variables. The DFXML format supports file system metadata not represented in the NSRL RDS or SWID formats, and is the base processing format used in the Diskprint analysis workflow, especially for file system differencing to determine new files. DFXML is provided here for some process provenance reporting (such as recording supporting library and script versions), and for reporting file modification times that were observed to appear consistently, which can indicate a file being copied from a server or installation media.

Further references:

Please contact nsrl@nist.gov with questions or feedback about any of these data sets.

More information about the SWID tag format can be found in the ISO specification or NIST's implementation guidance, respectively available at:
https://www.iso.org/standard/65666.html https://dx.doi.org/10.6028/NIST.IR.8060

More information about the DFXML format can be found in a journal article available from Digital Investigation or the article's lead author's website, and in the XML schema that specifies the format's current state. These are respectively available at:
https://dx.doi.org/10.1016/j.diin.2011.11.002
https://simson.net/clips/academic/2012.DI.dfxml.pdf
https://github.com/dfxml-working-group/dfxml_schema