

CLAIMS

1. A computer implemented method for maintaining data processing system security, the method implemented by one or more computer processors comprising:

receiving a plurality of streams, each corresponding to a respective data file, into a data store;

for each of the corresponding respective data files:

calculating a respective checksum using at least part of a corresponding respective stream of the plurality of streams,

selecting respective corresponding metadata from a metadata storage unit by matching the respective calculated checksum with a checksum stored in correspondence with the metadata, and

retrieving respective one or more virus indicators from the selected respective corresponding metadata;

generating a prioritized list including the each corresponding respective data files by rating the priority of the each corresponding respective data files in order of a respective likelihood, determined by the respective one or more retrieved virus indicators, that the each of the data files contains malware;

conducting a scan of a first data file having a highest priority in the prioritized list before conducting a scan of another data file in the prioritized list;

after the conducting, receiving a new stream corresponding to a new data file;

retrieving new one or more virus indicators corresponding to the new data file from newly selected corresponding metadata from the metadata storage unit;

re-generating the prioritized list, including the each corresponding respective data file that have not been scanned and the new data file, by rating the priority of the each corresponding respective data files that have not been scanned and the new data file in order of a respective likelihood, determined by the respective one or more retrieved virus indicator and the new one or more virus indicators, that the each of the data files and the new data file contains malware;

scanning each file in the regenerated prioritized list in order, beginning with the highest prioritized file in the re-generated prioritized list; and

blocking user access to each file in the re-generated prioritized list until after the each file is scanned.

2. The computer implemented method according to claim 1, wherein each respective checksum is continuously calculated in parallel as each corresponding respective data file streams into the data store.

3. The computer implemented method according to claim 1, further comprising: updating, by the one or more computer processors, each respective corresponding metadata of each respective corresponding data file according to the scan of the each respective corresponding data file; and providing, by the one or more computer processors, each respective corresponding data file for further processing.

4. The computer implemented method according to claim 1, wherein the prioritised list is generated according to a correlation.

5. The computer implemented method according to claim 1, further comprising presenting, by the one or more computer processors, at least one of the respective data files to a cleansing routine.

6. The computer implemented method according to claim 1, further comprising updating, by the one or more computer processors, one of the checksums stored in correspondence with the metadata.

7. A computer program product for maintaining system security, the computer program product comprising one or more computer readable storage devices including non-transitory computer readable memory including stored program instructions thereon, the stored program instructions, when executed, causing the execution of steps comprising:

receiving a plurality of streams, each corresponding to a respective data file, into a data store;

for each of the corresponding respective data files;

calculating a respective checksum using at least part of a corresponding respective stream of the plurality of streams,

selecting respective corresponding metadata from a metadata storage unit by matching the respective calculated checksum with a checksum stored in correspondence with the metadata, and

retrieving respective one or more virus indicators from the selected respective corresponding metadata;

generating a prioritized list including the each corresponding respective data files by rating the priority of the each corresponding respective data files in order of a respective likelihood, determined by the respective one or more retrieved virus indicators, that the each of the data files contains malware;

conducting a scan of a first data file having a highest priority in the prioritized list before conducting a scan of another data file in the prioritized list;

after the conducting, receiving a new stream corresponding to a new data file;
 retrieving new one or more virus indicators corresponding to the new data file from newly selected corresponding metadata from the metadata storage unit;
 re-generating the prioritized list, including the each corresponding respective data files that have not been scanned and the new data file, by rating the priority of the each corresponding respective data files that have not been scanned and the new data file in order of a respective likelihood, determined by the respective one or more retrieved virus indicators and the new one or more virus indicators, that the each of the data files and the new data file contains malware;
 scanning each file in the re-generated prioritized list in order, beginning with the highest prioritized file in the re-generated prioritized list; and
 blocking user access to each file in the re-generated prioritized list until after the each file is scanned.

8. The computer program product according to claim 7, wherein each respective checksum is continuously calculated in parallel as each corresponding respective data file streams into the data store.

9. The computer program product according to claim 7, further comprising: program instructions to update each respective corresponding metadata of each respective corresponding data file according to the scan of the each respective corresponding data file; and provide each respective corresponding data file for further processing.

10. The computer program product according to claim 7, wherein the prioritized list is generated according to a correlation.

11. The computer program product according to claim 7, further comprising program instructions to present at least one of the respective data files to a cleansing routine.

12. The computer program product according to claim 7, further comprising program instructions to update one of the checksums stored in correspondence with the metadata.

13. A computer system for maintaining system security, the computer system comprising: one or more computer processors; and one or more computer readable storage devices including non-transitory

computer readable memory including, stored program instructions thereon for execution by the one or more computer processors, the stored program instructions comprising program instructions to:

receive a plurality of streams, each corresponding to a respective data file, into a data store;

for each of the corresponding respective data files;

calculate a respective checksum using at least part of a corresponding respective stream of the plurality of streams,

select respective corresponding metadata from a metadata storage unit by matching the respective calculated checksum with a checksum stored in correspondence with the metadata, and

retrieve respective one or more virus indicators from the selected respective corresponding metadata;

generate a prioritized list including the each corresponding respective data files by rating the priority of the each corresponding respective data files in order of a respective likelihood, determined by the respective one or more retrieved virus indicators, that the each of the data files contains malware;

conduct a scan of a first data file having a highest priority in the prioritized list before conducting a scan of another data file in the prioritized list;

after the conducting, receive a new stream corresponding to a new data file;

retrieve new one more virus indicators corresponding to the new data file from newly selected corresponding metadata from the metadata storage unit;

re-generate the prioritized list, including the each corresponding respective data files that have not been scanned and the new data file, by rating the priority of the each corresponding respective data files that have not been scanned and the new data file in order of a respective likelihood, determined by the respective one or more retrieved virus indicators and the new one or more virus indicators, that the each of the data files and the new data file contains malware;

scan each file in the re-generated prioritized list in order, beginning with the highest prioritized file in the re-generated prioritized list; and

block user access to each file in the re-generated prioritized list until after the each file is scanned.

14. The computer system according to claim 13, wherein each respective checksum is continuously calculated in parallel as each corresponding respective data file streams into the data store.

15. The computer system according to claim 13, further comprising: program instructions to update each respective corresponding metadata of each respective corresponding data file according to the scan of

the each respective corresponding data file; and provide each respective corresponding data file for further processing.

16. The computer system according to claim 13, wherein the prioritized list is generated according to a correlation.

17. The computer system according to claim 13, further comprising program instructions to present at least one of the respective data files to a cleansing routine.

18. The computer system according to claim 13, further comprising program instructions to update one of the checksums stored in correspondence with the metadata.