

ABSTRACT

Data is efficiently extracted based on a communication risk.

5       ~~A control unit (11) determines risk levels of pseudo session data (dp) and session data segment (d3) and collects file information from the pseudo session data (dp) and the session data segment (d3). In this case, because no file is included in the session data segment (d3) and a file is included in the pseudo session data (dp), file information is collected from the file included in the pseudo session data (dp). In addition, an a.exe file is included in the pseudo session data (dp) and is associated with a schtasks command associated with 3 as the risk level. Thus, the control unit (11) collects, as the file information, meta information, a file hash value, and a file main body from the a.exe file.~~A control unit generates a group of data segments by dividing the communication data into sessions, extracts, from the group of data segments, data segments which have identical identification information and whose session interval is equal to or less than a threshold, generates linked data by linking the extracted data segments, determines a risk on communication based on certain information included in the linked data, and collects file information based on the risk from a file included in the linked data.

10

15

20

25

FIG. 1