

CLAIMS

1. A computer-implemented method for searching similar malwares based on their dynamic analysis results, the method comprising:

 a preparatory step, comprising:

analyzing a plurality of malwares in an isolated environment;

recording actions executed by each of the plurality of malwares in a plurality of behavioral reports;

separating the plurality of behavioral reports accumulated so that at least one group contains reports comprising similar actions and related to one known malware family;

creating a feature vector based on the known malware family according to selected fields from the plurality of behavioral reports;

training a plurality of binary classifiers on the basis of each obtained feature vector; and

creating at least one ensemble of the binary classifiers based on the pretrained binary classifiers;

 an operating step, comprising:

analyzing at least one malware;

recording actions executed by at least one malware in at least one behavioral report;

starting the ensemble of the binary classifiers trained at the preparatory step; and

calculating, by the ensemble of the binary classifiers, a resulting estimate that the analyzed malware belongs to at least one known malware family.

2. The method of claim 1, wherein analyzing malwares in the isolated environment comprises recording at least the following:

processes started by each of the plurality of malwares, each process starting parameters;

files created by each process;

mutexes created by each process;

registry keys created by each process;

virtual machine screen video during each of the plurality of malwares running;

network activity; and

harmfulness markers.

3. The method of claim 1, wherein the plurality of behavioral reports are reports in JSON format comprising at least the following fields regarding malware analysis: IP addresses, registry keys, mutexes, user agents, markers, processes and running programs, request type and domain, files.

4. The method of claim 1, wherein the resulting estimate of the ensemble of the classifiers is the resulting estimate from -1 to 1, where -1 means that at least one malware does not belong to the specified family at all, and 1 means that at least one malware definitely belongs to the specified family.
5. The method of claim 1, wherein at least one ensemble of the classifiers is configured to detect reports related to at least one malware group.
6. The method of claim 1, wherein the feature vector is composed based on attributes collected using regular expressions from the behavioral report.
7. The method of claim 1, wherein the separating the plurality of behavioral reports is executed on the basis of a database of tagged known malicious files using at least ssdeep, YARA, IDS, behavioral signatures, specific behavioral features, which could be related to the specific malware family.
8. A system for searching similar malwares based on their dynamic analysis results, the system comprising:
 - a computing device configured to implement the above method according to claims 1-7;
 - a recording module configured to record actions of at least one malware in at least one of the plurality of behavioral reports; and
 - an output module.