Monthly Research

An Example of Antivirus Detection Rates and Similarity of Undetected Malware

FFRI,Inc
http://www.ffri.jp

Ver 2.00.01

Agenda

- Background and purpose
- Surveyed Malware
- About FFRI Dataset
- An example of antivirus detection rates
- Considerations of the detection rates
- Similarity of undetected malware by fuzzy hashing
- Summary

Background and purpose

- The executive of antivirus vendor said "antivirus software is dead" in May 2014.
  It became a conversation topic.

- In this slides, we show an example of antivirus detection rates.

- We investigated similarity of undetected malware by fuzzy hashing to improve detection rate.

Surveyed Malware

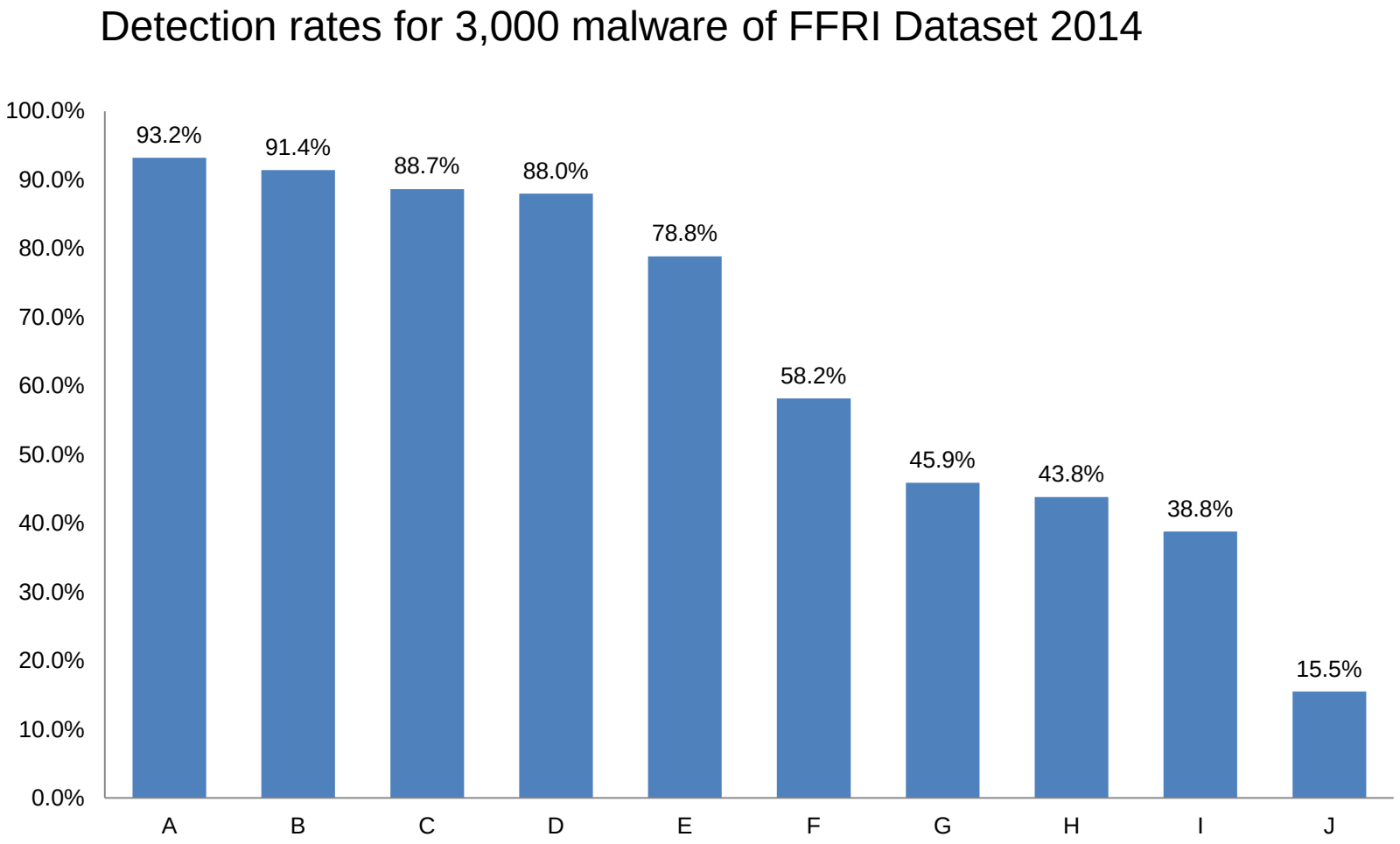
- Property
  - Known malware
  - Those are not targeted malware which is hard to detect in general.
- Collection period
  - From January 2014 to April 2014
- Number of malware
  - 3,000
- Investigation period of detection rates
  - April 24, 2014 and July 7, 2014

About FFRI Dataset

- We have provided the dataset for anti-malware research.
  - It is available in the MWS (Anti-Malware Workshop).
    - http://www.iwsec.org/mws/2014/

- Overview of the FFRI Dataset 2014
  - Dynamic analysis log (Malware behavior information) of the foregoing malware.
  - It was generated by Cuckoo Sandbox and FFR yarai analyzer Professional.
  - See below for more information.
    - http://www.iwsec.org/mws/2014/files/FFRI_Dataset_2014.pdf

An example of antivirus detection rates



Considerations of the detection rates

- There were unexpected differences in detection rates.
- The detection rates did not change almost, after 2 months.
- Generic detection (pattern matching) was better than reputation-based detection.
- Free products was low performance.
- Attention points
  - The result is only by static detecting.
  - Therefore, it is not comprehensive evaluation.

Similarity of malware by fuzzy hashing

- Fuzzy hashing is a kind of technology to identify similar files.
  - Refer to Monthly Research in March 2014 for more information.

- We investigated following items by fuzzy hashing.
  - How many similar pairs are there in undetected malware group?
  - How many undetected malware which are similar to detected malware are there?

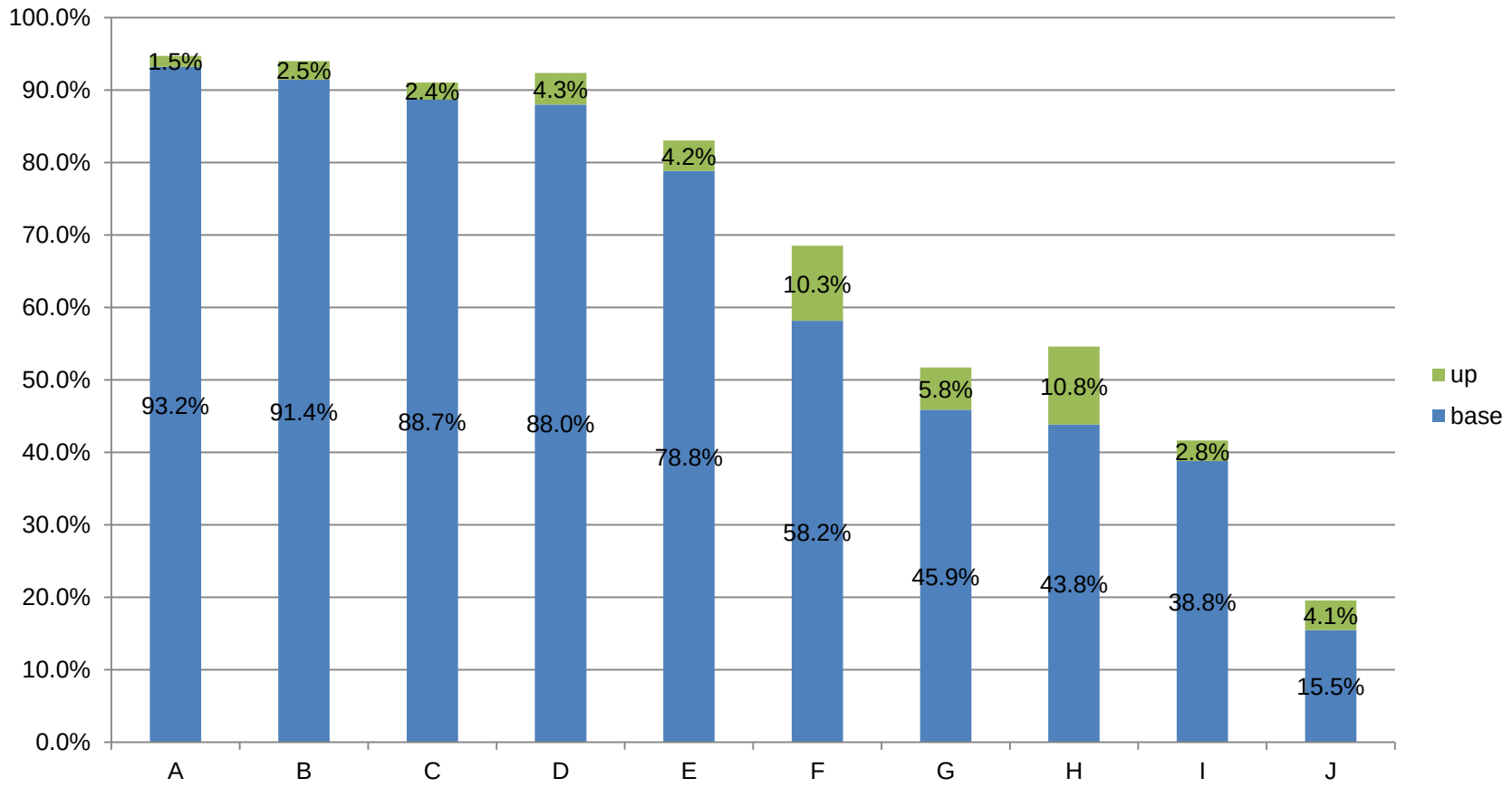
- We used the sdhash is a fuzzy hashing tool.
  - http://sdhash.org
  - Definition of similar pair is sdhash score over 21.

Result: Similarity of Undetected Malware

Vendor	Undetected	Similarly avg.	Similar to detected
A	197	2.89	44
B	253	8.11	75
C	340	2.29	71
D	360	3.72	130
E	628	4.33	124
F	1250	6.51	309
G	1620	16.01	175
H	1670	11.30	320
I	1835	15.95	85
J	2580	14.24	106

- Similarly avg. - Average number of similar malware when sampling one undetected malware.

Percentage of undetected malware which are similar to detected malware



Considerations

- Most vendors possibly require 100 to 200 additional patterns to detect undetected malware.

- There are not many undetected malware which are similar to detected malware.
  - It is going to improve detection rate of 10.8% in the best, if detecting the undetected malware which are similar to detected malware.
  - However, these are improvements of less than 5% for most vendors.

Summary

- There are great differences between static detection rates of antivirus in the FFRI Dataset 2014.
- Most vendors possibly require 100 to 200 additional patterns to detect undetected malware.
- There are not many undetected malware which are similar to detected malware.
- Behavior detection is required because static detection is reaching the limit.

Contact Information

E-Mail ： research-feedback@ffri.jp
Twitter： @FFRI_Research