Microsoft

Independent Tests of Anti-Virus Software

Does prevalence matter?

Ranking antimalware products by potential victim impact

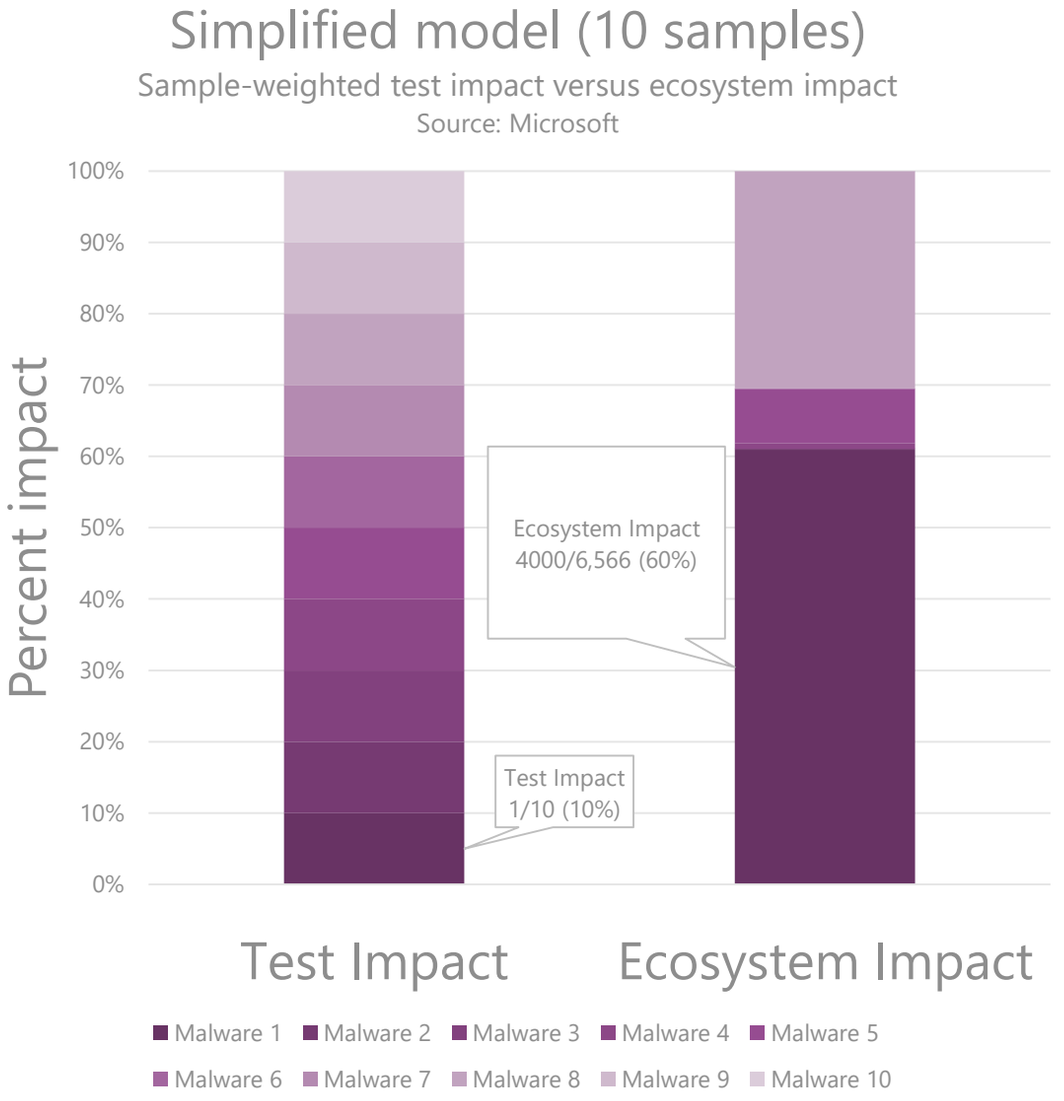
Peter Stelzhammer, AV-Comparatives
Holly Stewart, Microsoft

Overview

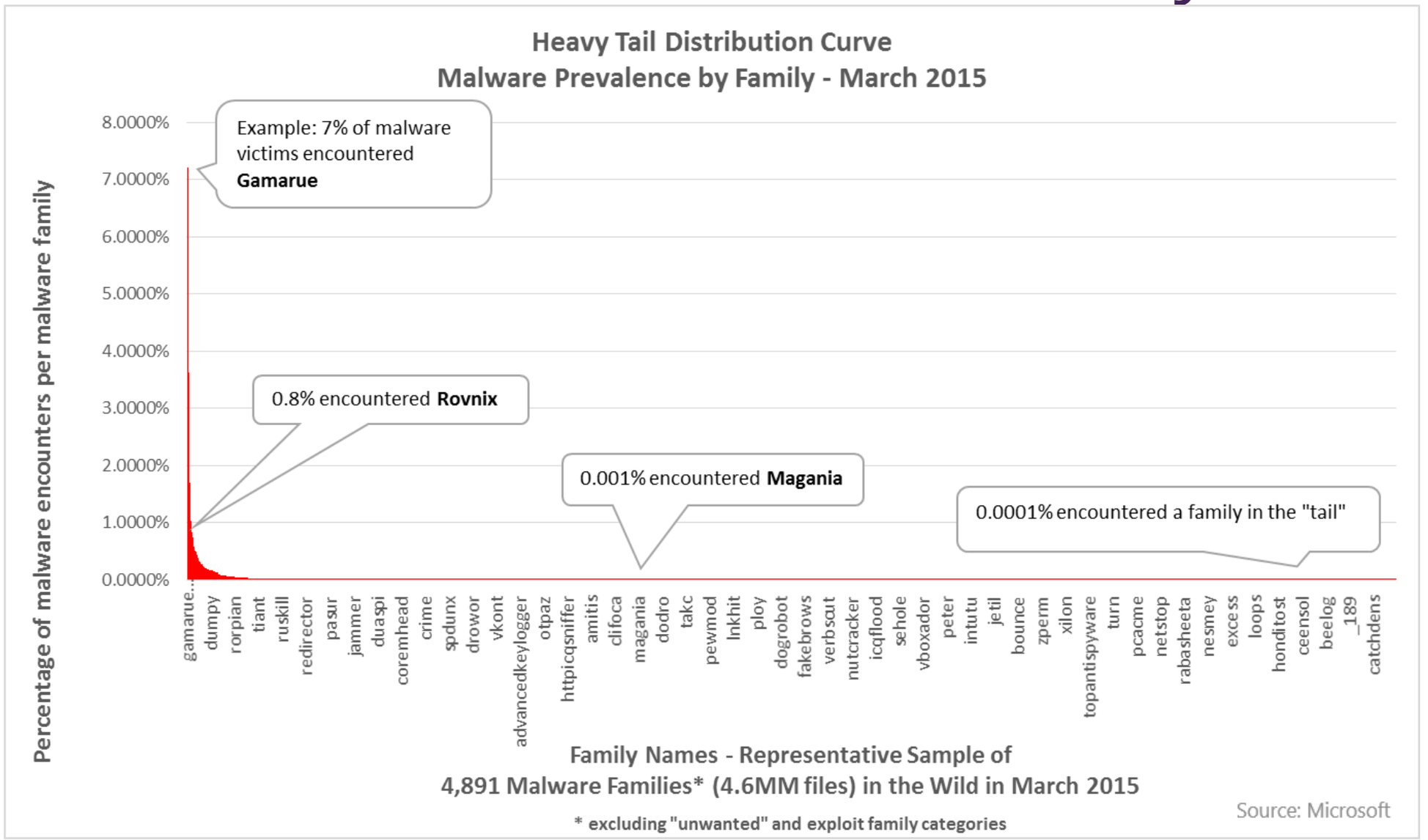
Today's scoring model compared to the ecosystem impact

Traditional tests count misses equally

Actual customer experience is different – some malware affects more people than others



Detailed look at the ecosystem



Challenges & tester constraints

Files in the test set should be…

Indisputable

No unwanted software, adware, etc.

PE (portable executable) files

Last month, PEs represented 64% of all malware Microsoft customers encounter.
Other file types include exploits, documents, malicious .lnk files, etc.

Recently discovered

PE files seen in the past 30 days represented 23%

Obtainable

Not all files are easy to obtain. Last month, new PE files obtained by Microsoft represented 4% of all files encountered.

Models

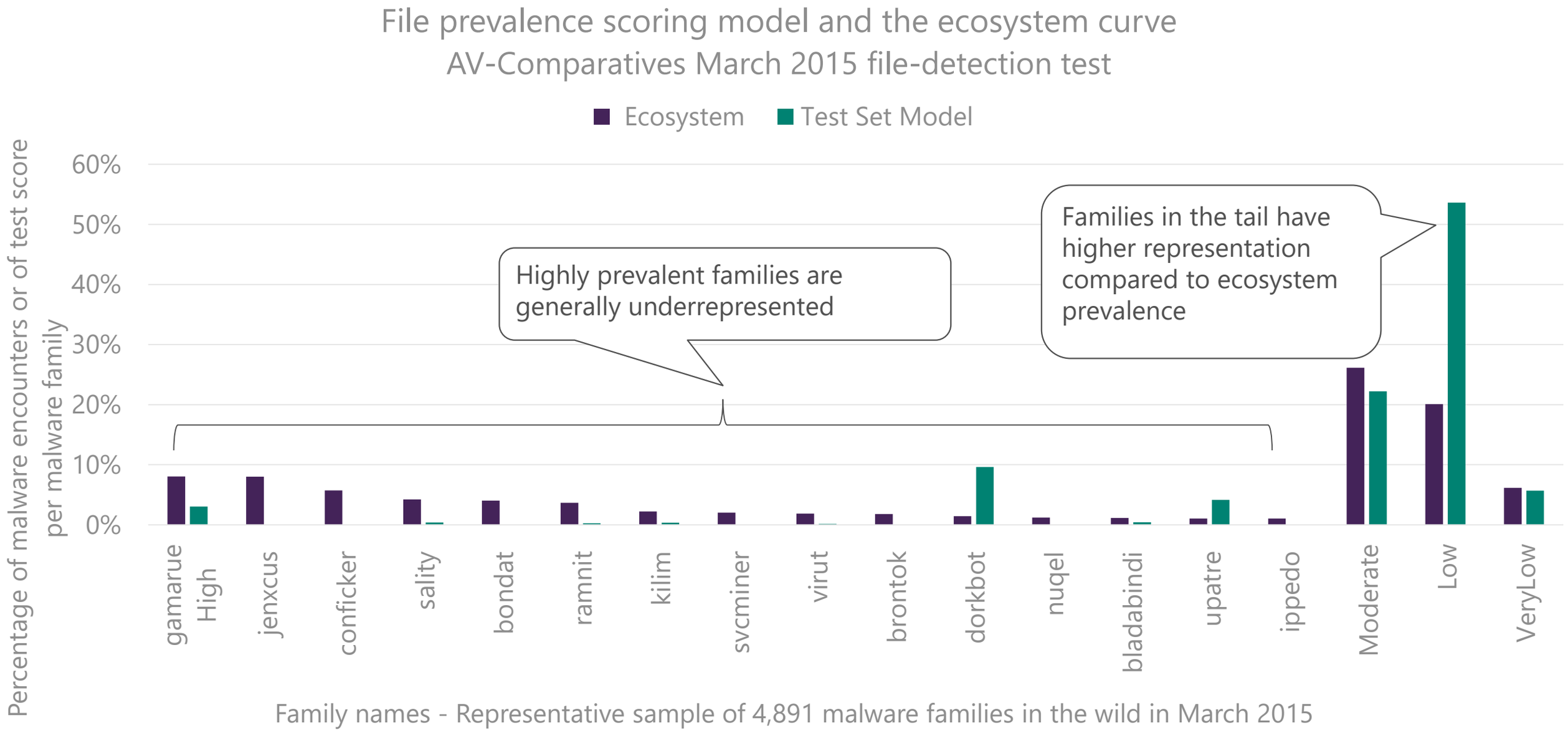
File prevalence

Definition: **Prevalence** is the # of distinct computers affected by a malicious file, malicious malware family or category of malware

$$\text{File prevalence weighted test score} = \frac{\text{prevalence of detected files}}{\text{prevalence of all files in test}}$$

Issues: Prolific, highly polymorphic families are underrepresented

File prevalence



File and family prevalence (2 models)

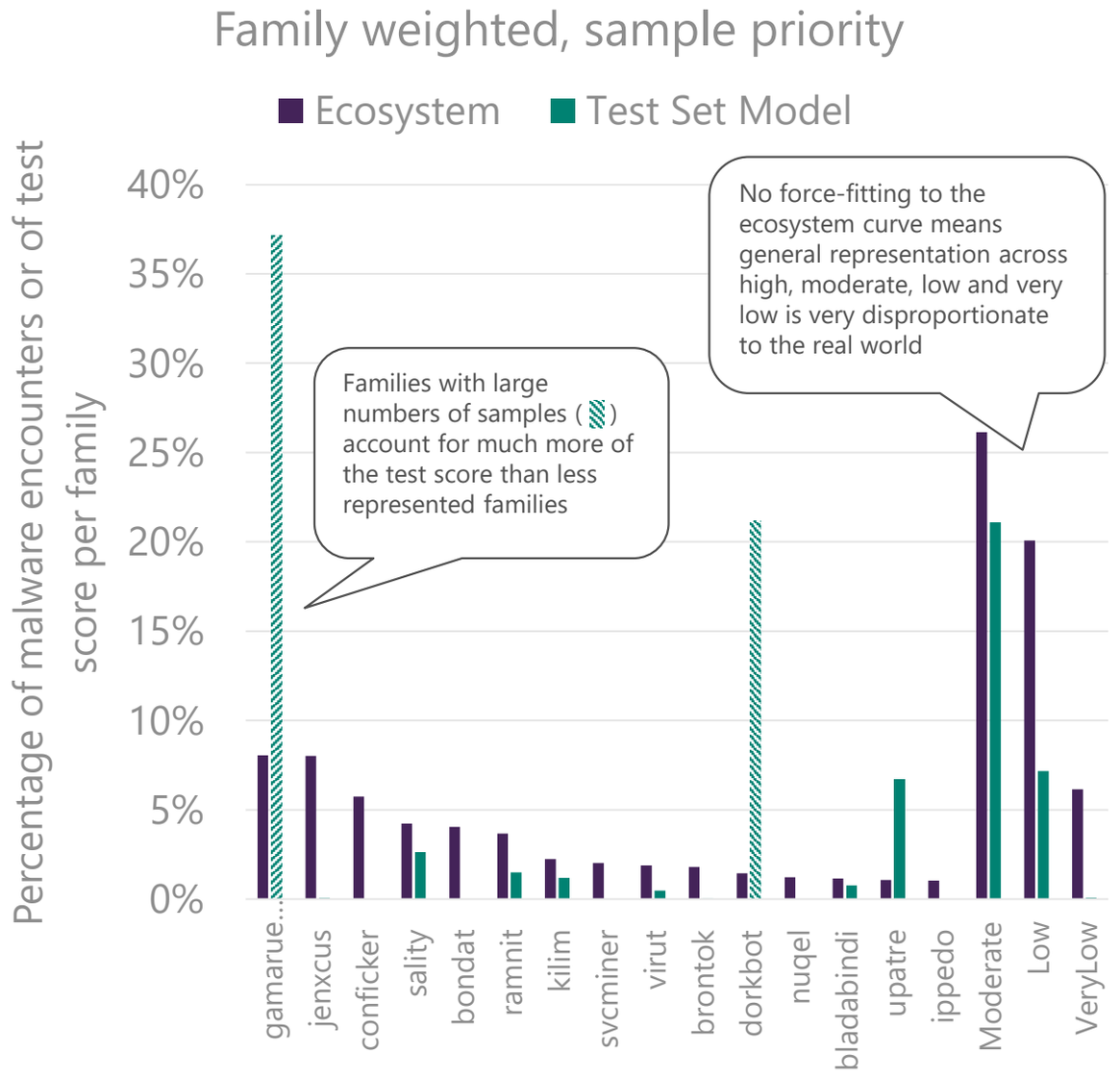
Weight the sample by file prevalence and also family prevalence

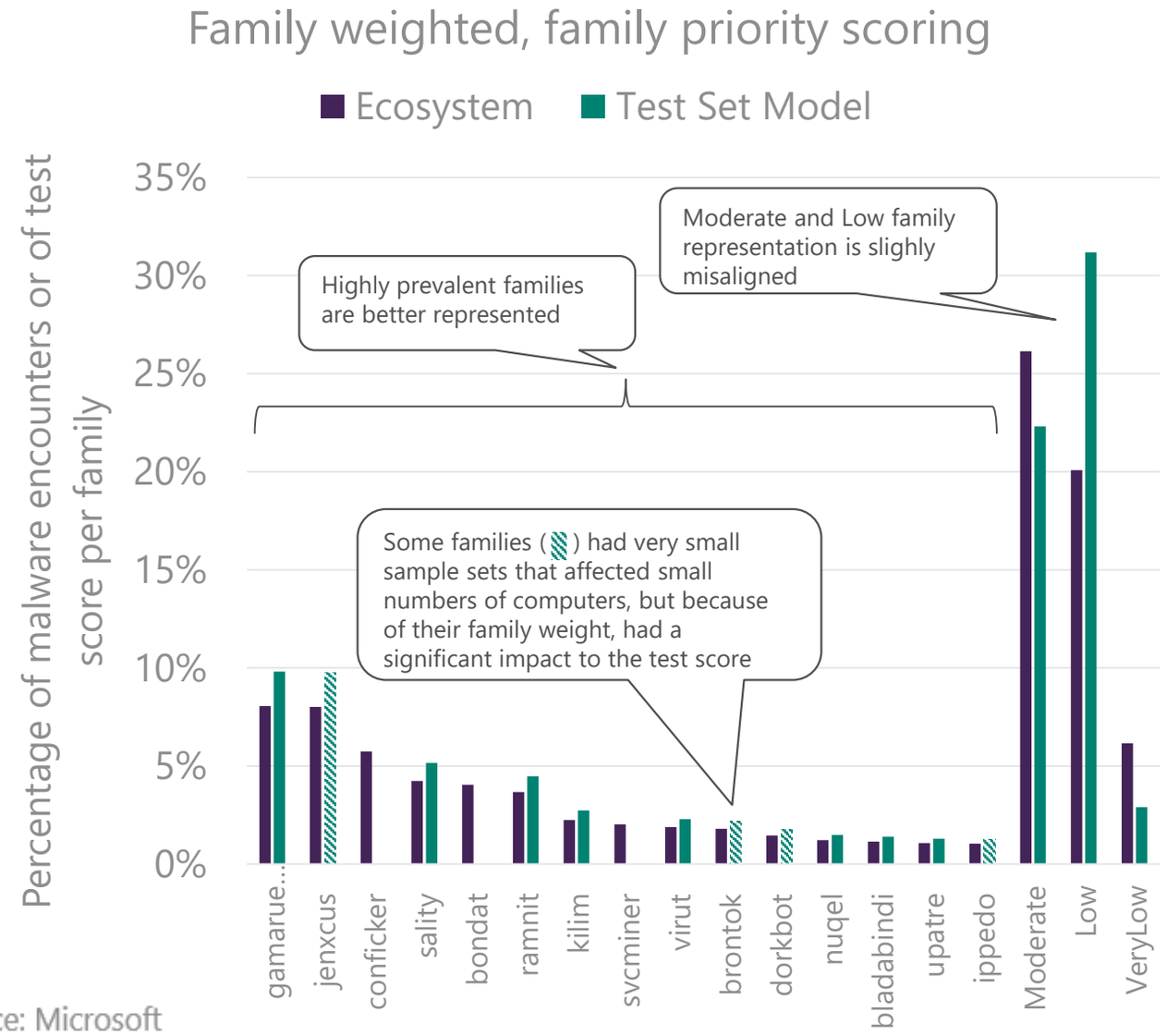
Example: A Gamarue file affecting 10 computers is modified by the family prevalence of 20%. (Whereas a smaller family with a sample affecting 10 computers would be modified by a smaller increase, say .01%)

Equate all samples of a particular family in the test to the prevalence of the family in the ecosystem

Example: If Gamarue is 20% of the ecosystem, then the sum of Gamarue files in the test equate to 20% of score

File and family prevalence (2 models)





Source: Microsoft

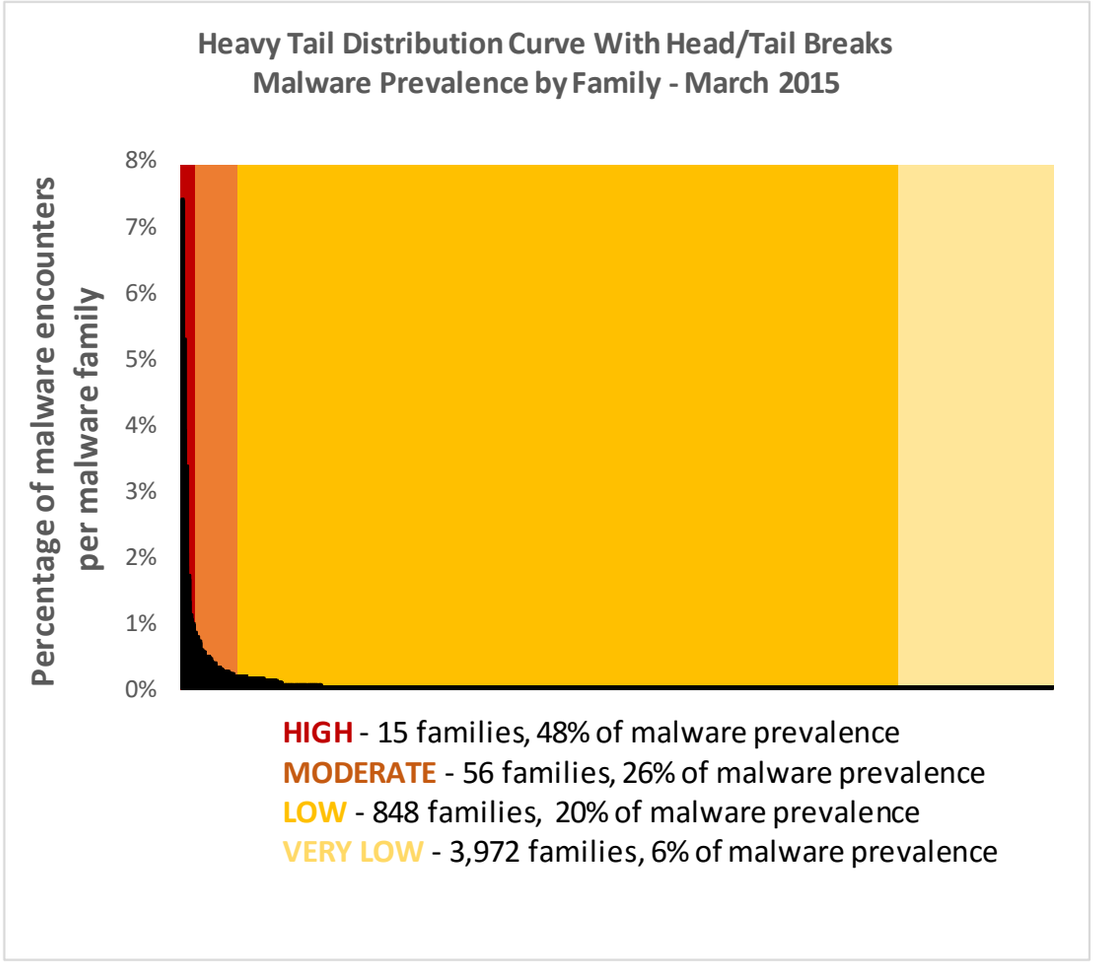
File, family, and family partition

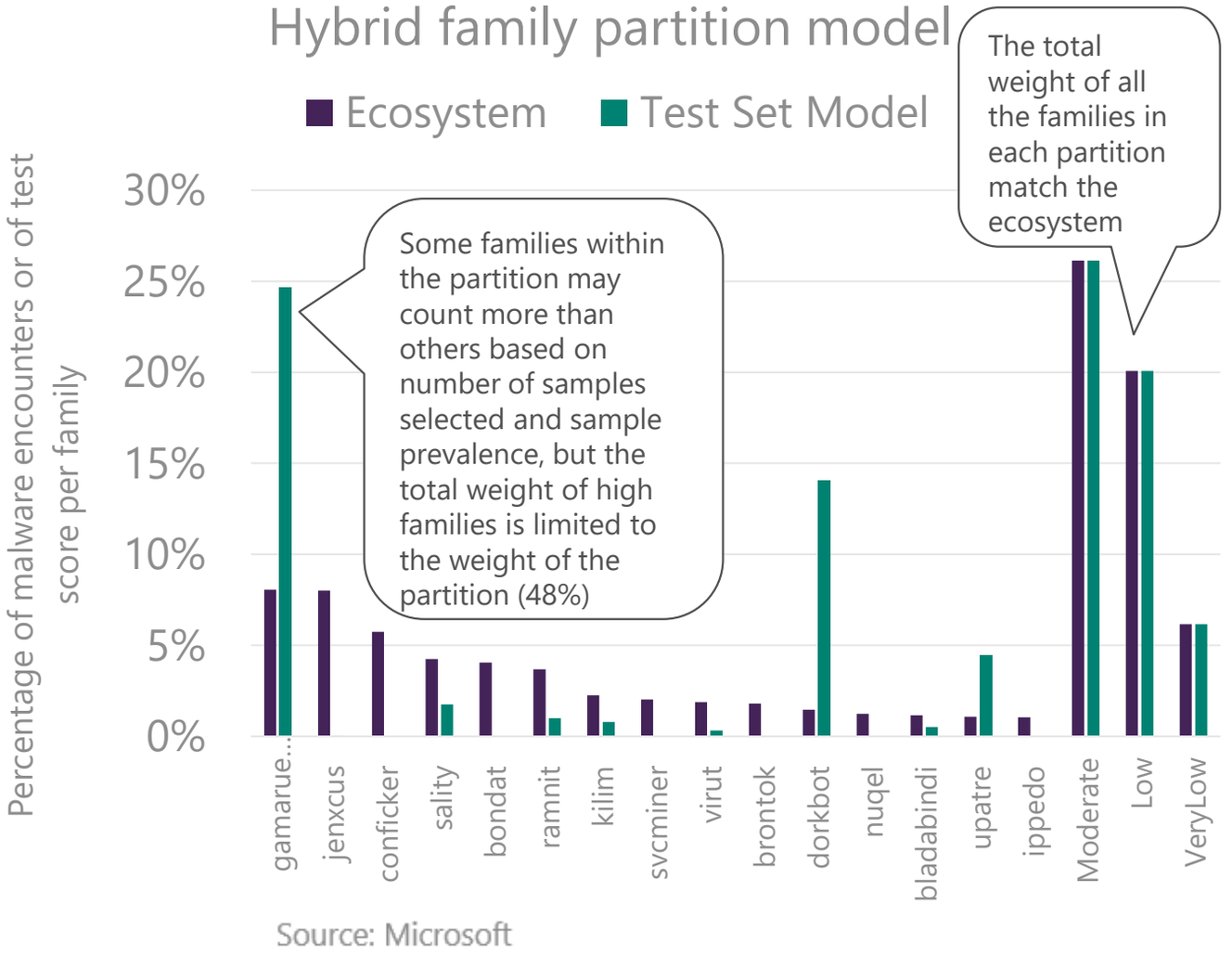
Allows file prevalence to be weighted by family prevalence, but force fits the test set to the ecosystem by partition rather than family

Benefits: Ensures the test set to match the ecosystem curve. Doesn't require the tester to have the "perfect" test set to represent all families

Drawbacks: Complicated to calculate and explain!

File, family, and family partition



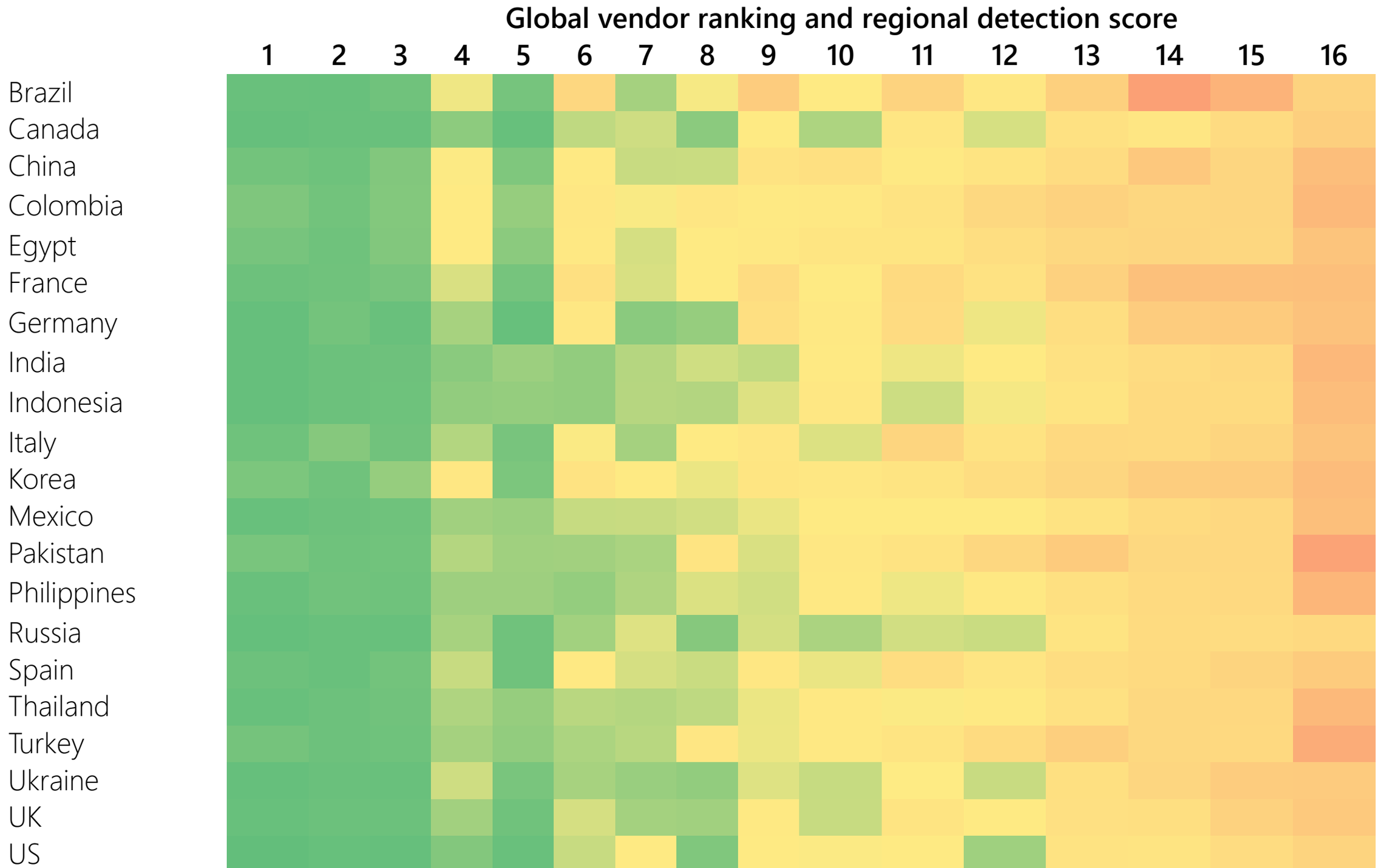


Results

Traditional vs prevalence-weighted

Vendor ranking - Traditional model	Vendor ranking - Prevalence model	Movement
1	1	-
2	2	-
3	5	(2)
4	8	(4)
5	3	2
6	7	(1)
7	11	(4)
8	4	4
9	10	(1)
10	6	4
11	9	2
12	14	(2)
13	12	1
14	17	(3)
15	13	2
16	15	1
17	16	1

	Traditional	Prevalence
Highest score:	99.96%	99.99%
Lowest score:	86.26%	98.83%


Global vendor ranking and regional detection score
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
Brazil
Canada
China
Colombia
Egypt
France
Germany
India
Indonesia
Italy
Korea
Mexico
Pakistan
Philippines
Russia
Spain
Thailand
Turkey
Ukraine
UK
US

Going forward

Lessons learned

It's nearly impossible for a traditional scoring model to represent the real world

Building one that does is complicated

Telemetry on global and local family and file prevalence would make the prevalence-weighted model more relevant

Call to action

AMTSO Realtime Threat List:

Support more data types (distinct machines, family prevalence, common timeframes, and locality)

Vendors:

To increase accuracy, share prevalence data on files, families and locality

High-quality input required (no junk)

Questions?

Microsoft

AV comparatives

Independent Tests of Anti-Virus Software