Comparing Detection Capabilities of AntiVirus Products: an Empirical Study with Different Versions of Products From the Same Vendors

Areej Algaith, Ilir Gashi
Center for Software Reliability
City University London
London, UK
{areej.algaith.1, ilir.gashi.1}@city.ac.uk

Bertrand Sobesto, Michel Cukier
A. James Clark School of Engineering
University of Maryland, College Park
MD, USA
{bsobesto, mcukier}@umd.edu

Selman Haxhijaha, Gazmend Bajrami
University for Business and Technology
Pristina, Kosovo
{selmanh, gazmend_bajrami}@hotmail.com

***Abstract* — In this paper we report results of an empirical analysis of the detection capabilities of 9 AntiVirus (AV) products when they were subjected to 3605 malware samples collected on an experimental network over a period of 31 days in November-December 2013. We compared the detection capabilities of the version of the AV products that the vendors make available for free in VirusTotal versus the full capability products that they make available via their own website. The analysis has been done using externally observable properties of the AV products: namely whether they detect a given malware. The paper reports extensive analysis of the results. A surprising finding of our study was that only one of the vendors had a full capability version which detected all the malware that their VirusTotal version could detect.**

Keywords— AntiVirus products; security assessment; quantitative assessment; empirical studies

I. INTRODUCTION

AntiVirus (AV) products are some of the most widely used security protection systems. They are usually deployed as the last line of defence on desktop, laptop, tablet and smartphone devices for both home and business use. Studies that compare their detection capabilities are widely available[1].

VirusTotal[2] is an online service that hosts (at the time of writing) 53 signature-based detection engines from different AV vendors. It is a service that is widely used by both academia and industry to submit and inspect malware samples. It also provides an Application Programming Interface (API) through which multiple malware samples can be submitted.

In this paper we report results of an empirical analysis of the detection capabilities of 9 AV products when they were subjected to 3605 malware samples collected on an experimental network over a period of 31 days. The results included a comprehensive analysis of the dataset along three dimensions (AV, malware, dates). We compared the detection capabilities of the version of the AV products that the vendors make available for free in VirusTotal versus the full capability products that they make available via their own website (for these versions of their products they usually require users, especially for business use, to pay a fee). The whole analysis has been done using externally observable properties of the AV products: namely whether they detect a given malware.

The malware used to send to the AV products have been collected on an experimental infrastructure which consists of a honeypot network distributed over several different countries. These honeypots collect malware, which we then send for inspection to VirusTotal. Each AV product either detects a malware or fails to detect. We then continued sending the collected malware to the signature-based detection engines in VirusTotal over a period spanning 11-November-2013 to 11-December- 2013 to observe the detection capabilities of the AV products in VirusTotal over time. Additionally, in the experimental infrastructure, we also deployed nine AV products which we downloaded from the websites of the vendors. We then sent the same malware on the same dates as for VirusTotal to these AV products. This allowed us to then compare the detection capabilities of two AV versions from the *same* nine vendors and also across these nine *different* vendors.

The main motivation and purpose for the research we report in this paper is to compare the detection rates of full capability AV products versus the AV version that the vendors make available for free via VirusTotal. By doing this analysis we provide results which help other security researchers (from academia or industry), who may have results from experimentations with VirusTotal alone, to compare and improve their estimates. VirusTotal state the following on their FAQ [3] : "VirusTotal's antivirus engines are commandline versions, so depending on the product, they will not behave exactly the same as the desktop versions". With our research we hope to provide some comparative statistics on the behaviour of full capability desktop versions and their VirusTotal versions for the given environment where we run them. We sent the malware to these AV products every day for a month, hence our analysis also allowed us to present results on how the detection capabilities of these versions evolved over time and whether there were differences in the evolution between the versions that

[1] av-comparatives.org/, av-test.org/, virusbtn.com/index
[2] https://www.virustotal.com/
[3] https://www.virustotal.com/en/faq/#statistics

the vendors make available in VirusTotal versus the ones they make available directly from their websites.

This paper is structured as follows: Section II contains descriptions of related work; Section III describes the experimental architecture that we have deployed to collect the results; Section IV presents the results of the empirical analysis we have performed of the detection capabilities of two versions of AV products from nine vendors; Section V provides a discussion of our results and finally Section VI presents conclusions and future work.

II. Related work

As we stated previously, there are numerous sites that report rankings and comparisons of AVs (see footnote 1 for links to these sites). An interesting analysis of "at risk time" for single AV products is given in [1]. In this paper the authors analyzed how long it takes for different AV vendors to detect a malware.

There have also been studies to assess the benefits in improved malware detection from using more than one diverse AV product. An initial implementation of an architecture called Cloud-AV, which utilizes multiple diverse AV products to detect the malware was given in [2]. The authors in [2] also describe an empirical analysis of the benefits of diversity based on the deployment of Cloud-AV at the University of Michigan network. Our own research on results from large-scale studies on the detection capabilities of diverse AV products, using VirusTotal versions of these AV products, are reported in [3],[4].

Over the past several years, researchers and practitioners have used honeypots to learn about attacks and attackers. These systems can be categorized as security tools whose value lies in being probed, attacked, or compromised [5]. These carefully monitored systems allow security researchers to attract hackers, analyze their actions and profile them [6].

Honeypot systems can be found at different scales: from a single host to more complex honeypot networks. These networks, also called honeynets, can be deployed on a few IP addresses within a local network. The project Leurre.com [7], SGNET [8],[9] and the honeynet initiative from CAIDA [10] are examples of distributed honeypot networks in different locations.

III. Experimental architecture

Dionaea[4], a low-interaction honeypot tool that emulates common Internet services, has been deployed on a distributed honeypot architecture to collect the malware analyzed in this study.

1136 public IP addresses have been used to implement Dionaea. These IP addresses are distributed across six different locations: France, Germany, Morocco and the USA.

However, this study does not intend to compare the malware collected on the different networks or locations. The goal of this paper is to study the detection capabilities for a malware set. The subnets do not have the same size and their configuration differs from one network to another. Note that none of the networks have the same security policy. Besides, these networks are not protected the same way.

The implemented configuration of Dionaea exposes different Internet services and protocols. For each of these services and protocols, Dionaea emulates vulnerabilities used to trap malware attempting to exploit them. Because of the nature of the vulnerabilities and protocols emulated, we mainly collect Windows Portable Executable (PE) files[5].

Binary files can be captured in different ways and they can have different formats. Only Microsoft Windows PE files were kept for this study. This format is easily identifiable and executable on any Windows operating system.

Once a day, a Perl script fetches the malware and the information relative to their capture (stored in a SQLite database) from the different Linux virtual machines. This script then submits to VirusTotal[6] the entire malware repository for analysis. When using the submission API, VirusTotal returns a scan key for each malware sample submitted. This scan key is composed of the binary's SHA1 hash and the submission timestamp. The Perl program stores the different scan keys returned by the website to later retrieve the analysis reports.

An additional Perl script is executed once all the malware has been submitted. This program uses the previously stored scan keys to fetch the analysis reports for all the malware in the repository.

Everyday a new database entry is created for each malware. This entry contains the information related to the VirusTotal submission. The AV product's names, versions and the malware signature names are also uploaded in various tables and linked together with the file submission.

We also deployed 9 AV products (AntiVir, AVG, Comodo, F-Secure, Kaspersky, McAfee, Microsoft, Sophos and Symantec) and sent the same malware samples we collected in our experimental architecture to these AV products on the same dates as when we sent them to VirusTotal. We chose these AV products because:

- they represent some of the most widely deployed AV products on the market;
- we have experience of using them in the past.

The rest of the analysis presented in this paper concerns the comparison of detection capabilities of the version of the AV product that these nine vendors have on VirusTotal and the versions that we deployed in our setup after downloading a full capability version of the AV product from the vendors' website.

IV. Results

A. Summary Results

We start our analysis with describing some initial statistics of the data. As previously mentioned, our empirical analysis is with 2 versions (VirusTotal version and full capability version) of AV products of 9 different vendors when they were subjected to 3605 malware samples collected in our experimental infrastructure, over a 31 day period (11-November-2013 – 11-December-2013). Hence the inputs to our empirical analysis

[4] http://dionaea.carnivore.it

[5] http://download.microsoft.com/download/9/c/5/9c5b2167-8017-4bae-9fde-d599bac8184a/pecoff_v8.docx

[6] VirusTotal is a web service providing online malware analysis based on several AV products

consist of a series of *triplets* {AV_{i*2}, $Malware_j$, Day_k}: a given malware *j* is inspected on a given date *k* by two versions of a given AV *i*. For each of these triplets we observe a *detection* (stored as 0, or no failure), or *no detection* (stored as 1, or *failure*). For those triplets where we see a detection we also store the labels which a given AV version assigns to a given malware on a given date.

We send malware to the two AV versions every day from the first date a malware is observed in the honeypots in our infrastructure until the last day of the observation period. However, the total number of triplets we observed was less than 3605 Malware * 31 Days * 9 AVs * 2 Versions. This is because:

- Each day we observed new malware in the infrastructure and we could not send to VirusTotal a malware to be inspected by an AV product from an earlier date.
- Some of the AV products in VirusTotal are not responsive on certain dates for certain malware and hence we have no results for them.

In this analysis we use all the triplets {AV_{i*2}, $Malware_j$, Day_k} for which we had observations from both versions of a given vendor allowing us to do a like-for-like comparison. Overall we had 958,972 * 2 such triplets, as shown in Table I.

B. *Comparison of the Detection Capabilities of the two Versions from Each Vendor*

Table II shows the counts of demands (detected and undetected) for the full capability and VirusTotal versions of the nine vendor products in our study. A demand is a pair {$Malware_j$, Day_k} which links a given malware *j* and the date *k* to a given version of an AV that inspected it. The total number of demands for the vendors are comparable, apart from F-Secure for which did not get responses for several days in VirusTotal.

A surprising first observation is that for seven out of nine of these vendors, the version of their products that they had available in VirusTotal had a better detection rate compared with their full capability products (for each vendor we have highlighted in green which version gives the better detection rate).

TABLE I. COUNTS OF DETECTIONS AND NON-DETECTIONS

Detection/Failure (DF)	Full Capability (FC)	Virus Total (VT)
DF=0 - No failure: Detection	944,718	946,375
DF=1 - Failure: No Detection	14,254	12,597
Total	958,972	958,972

Additionally, we were also interested in finding out what were the detection rates for the AV products when they first had to inspect a newly observed malware in our honyepots. Table II shows these results. The ordering is similar to Table III with the exception of Comodo which fails to detect more malware the first time it encounters in the full capability version compared with the VirusTotal version.

TABLE II. DETECTIONS AND NON-DETECTIONS FOR ALL DEMANDS

Antivirus Name	DF_{FC}=0	DF_{FC}=1	FR_{FC}	DF_{VT}=0	DF_{VT}=1	FR_{VT}
AntiVir	108,141	455	0.004190	108,171	425	0.003914
AVG	108,051	61	0.000564	106,235	1877	0.017362
Comodo	108,423	148	0.001363	108,101	470	0.004329
F-Secure	91,339	1436	0.015478	91,383	1392	0.015004
Kaspersky	106,348	2146	0.019780	106,380	2114	0.019485
McAfee	103,969	4463	0.041159	106,392	2040	0.018814
Microsoft	105,817	2560	0.023621	105,974	2403	0.022173
Sophos	105,826	2612	0.024088	106,635	1803	0.016627
Symantec	106,804	373	0.003480	107,104	73	0.000681

FR_{FC} : Failure Rate of Full Capability Products
FR_{VT} : Failures Rate of Virus Total Products

TABLE III. DETECTIONS AND NON-DETECTIONS FOR THE FIRST INSPECTION OF A MALWARE BY AN AV VERSION IN OUR EXPERIMENT

Antivirus Name	DF_{FC}=0	DF_{FC}=1	DF_{VT}=0	DF_{VT}=1
AntiVir	3590	15	3591	14
AVG	3603	2	3543	62
Comodo	3600	5	3603	2
F-Secure	3549	56	3550	55
Kaspersky	3535	70	3536	69
McAfee	3410	195	3538	67
Microsoft	3526	79	3526	79
Sophos	3459	146	3546	59
Symantec	3535	70	3604	1

DF_{FC} : Detection Failures of Full Capability Products
DF_{VT} : Detection Failures of Virus Total Products

The two preceding tables give a good overview of the detection capabilities of the versions separately. We then checked in more detail which demands are being detected in one version but not the other (and vice versa).

Tables IV and V give these numbers for all demands (Table IV) and for the first inspection of a given malware by a given AV version (Table V).

TABLE IV. DETECTIONS AND NON-DETECTIONS FOR ALL DEMANDS ON BOTH VERSIONS

Antivirus Name	DF_{FC}=0 AND DF_{VT}=0	DF_{FC}=0 AND DF_{VT}=1	DF_{FC}=1 AND DF_{VT}=0	DF_{FC}=1 AND DF_{VT}=1
AntiVir	108,141	0	30	425
AVG	106,235	1816	0	61
Comodo	108,016	407	85	63
F-Secure	91,337	2	46	1390
Kaspersky	106,334	14	46	2100
McAfee	103,969	0	2423	2040
Microsoft	105,817	0	157	2403
Sophos	105,820	6	815	1797
Symantec	106,760	44	344	29

The total of the demands in the different columns of the two tables are as follows. First column: detected by both versions; second column: detected by the full capability product but not by VirusTotal; third column: detected by VirusTotal but not by the full capability product; and the fourth column: failed to be detected by both versions. For three products (AntiVir, McAfee and Sophos) we see no demands that have been detected by VirusTotal but not by the full capability products – for these products there seems to be no gain in detection capability from using the full capability product For AVG, we have the opposite observation: the full capability product detected everything that the signature based detection engine detects in VirusTotal and more. For the other five vendors, the detection capabilities of the

two versions of the product seem complementary: one version detected some demands that the other one cannot, and vice versa.

TABLE V. DETECTIONS AND NON-DETECTIONS FOR THE FIRST INSPECTION OF A MALWARE BY AN AV VERSION IN OUR EXPERIMENT – CATEGORISED BY COUNTS ON BOTH VERSIONS PER VENDOR

Antivirus Name	$DF_{FC}=0$ AND $DF_{VT}=0$	$DF_{FC}=0$ AND $DF_{VT}=1$	$DF_{FC}=1$ AND $DF_{VT}=0$	$DF_{FC}=1$ AND $DF_{VT}=1$
AntiVir	3590	0	1	14
AVG	3543	60	0	2
Comodo	3600	0	3	2
F-Secure	3550	0	1	54
Kaspersky	3535	0	1	69
McAfee	3410	0	128	67
Microsoft	3526	0	0	79
Sophos	3459	0	87	59
Symantec	3575	0	29	1

C. *Visualising the Dataset over the Three Dimensions (AV, Malware, Dates)*

Next, we investigated more closely the overall distributions of the detection rates to analyse any patterns or anomalies in the detection capabilities of the different vendors. Figures 1, 2 and 3 show this visualisation. Each of the figures represent three dimensional plots, with the *x* and *y* axes representing any two of the three dimensions of interest (Malware, AV or Date), and the *z* axis (given by the intensity of the colour) represents the proportion of demands of the remaining third dimension that have detection failures. We will use Figure 1(a) for explanation: the x-axis contains the dates (ordered from start to finish) of the collection period; the y-axis shows the malware (ordered by MD5 – same ordering preserved in parts (a) and (b) of the figure to make the visual comparison easier). A cell on the plot shows the proportion of full capability AV products that failed to detect a given malware on a given date. The colours of the cells represent the proportion of failures: white colour means none of the AV products failed to detect a given malware at a given date (i.e. they all detected the malware); black colour means missing data; the range from light green to dark red represents the failure rates from greater than 0 to 1 (in this case the failure rate is given as a proportion of AV versions that failed to detect a given malware at a given date). Figure 2(b) shows the same plot for the VirusTotal. The cells in Figures 2 and 3 show the proportion of dates (Figure 2) and malware (Figure 3) that failed to be detected for a given AV/Malware or AV/Date respectively. The colour encoding is the same. Full analysis results for Figure (1 and 2) are provided in a technical report [11].

Main observations from these plots:

- Figure 1: it is visually clear that the full capability versions have a higher failure rate (due to the greater prevalence of non-white cells in Figure 1 (a)). A number of malware have a high failure rate throughout the period for both setups (as is clear from the red lines that run across the figure).
- Figure 2: There is some visible diversity in the "difficulty" of the malware across different vendors: we have red lines that run across several AV products (maximum 6 AVs for full capability products; maximum 8 AVs for VirusTotal versions). The detection rates that we observed in Table II are confirmed in the figure: Comodo and AVG have a lower number of coloured cells in part (b) of the figure compared with (a); vice versa for the others.
- Figure 3: Only the VirusTotal version of Symantec has a perfect detection rate of all malware on a few dates of the experiment (as seen from the white gaps in figure 3(b) for Symantec). The rest of the versions all fail on at least one malware. None of the full capability versions of AVs had a perfect detection rate on all the malware on any dates of the collection period. Comodo has a few days in the experimental period with a high failure rate in VirusTotal (as can be seen from the red areas in the top part of figure 3(b) for Comodo). It is not clear why this is as VirusTotal is a black box for us. We can speculate that during this period Comodo was not updating its signature database in VirusTotal (or the update led to it failing to detect malware that it had detected in the past).

D. *Time Lag Analysis*

Apart from looking at the detection capabilities of the two versions of the AV vendors in our study, we also looked at which version first detected a malware and what is the timelag between detections of malware by the two versions. The results are given in Table VI.

TABLE VI. TIME LAG BETWEEN DETECTIONS BY THE TWO VERSIONS OF AN AV VENDOR

		AntiVir	AVG	Comodo	F-Secure	Kaspersky	McAfee	Microsoft	Sophos	Symantec
	A	14	2	2	54	69	67	79	59	1
	B	1	0	3	0	1	17	0	16	0
	C	0	60	0	0	0	0	0	0	0
Difference in Days	0	3590	3543	3600	3550	3535	3410	3526	3459	3575
	+1						76			27
	+2				1		9			1
	+3								36	
	+4								35	
	+5						1			
	+11						2			
	+14									1
	+17						2			
	+22						11			
	+23						10			

- **A :** Malware have not been detected from both versions
- **B :** Malware have been detected from VT version but not detected by FC
- **C :** Malware have been detected from FC version but not detected by VT
- **Difference in days**: number of days elapsed from VT detecting a malware until the FC version from the same vendor also detects it.

The first column (Difference in Days from 0 to +23) represents the timelag in days: a 0 value means that both versions detected a malware on the same day; a positive value gives the difference in days between the first date that a full capability version detected a malware and the first time it was detected by VirusTotal version (e.g. a value of +2 means the VirusTotal version detected the malware two days ahead of full capability); The subsequent columns then give the counts of malware for each vendor. We should be clear that this is the malware for which both AV versions of a given vendor eventually did detect the malware: what we are measuring here is the difference in the time it took each vendor to first detect it in our collection period. Most of the malware is either detected on the same date or VirusTotal detects them a day earlier: this might be because of the slight delay with which we send the

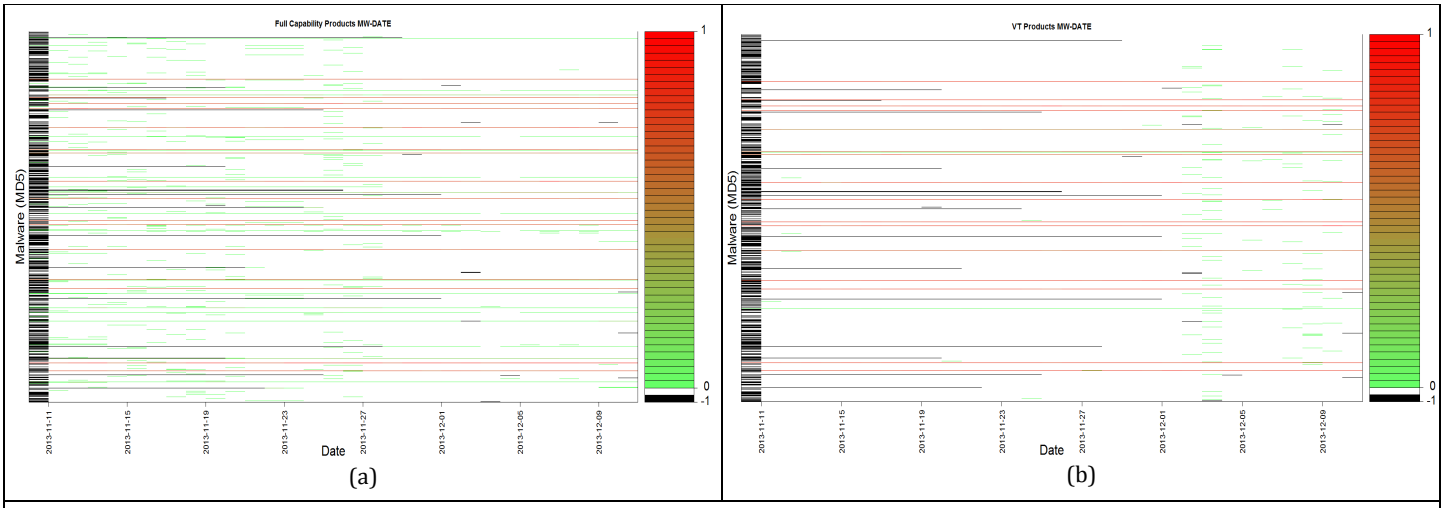

Fig. 1. Date (x-axis), Malware (y-axis) and the proportion of AVs (given by the intensity of the colour in the plot) that fail to detect a malware on a given date. Figure 1 (a) full capability versions; Figure 1(b) signature based detection engine as found in VirusTotal.

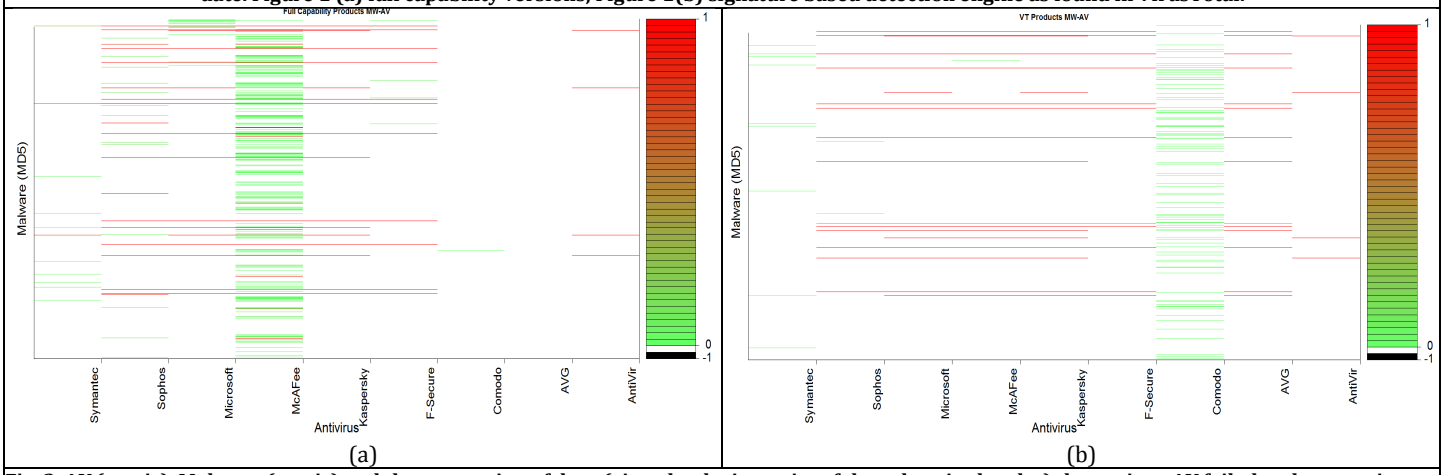

Fig. 2. AV (x-axis), Malware (y-axis) and the proportion of days (given by the intensity of the colour in the plot) that a given AV failed to detect a given Malware. Figure 2 (a) full capability versions; Figure 2(b) signature based detection engine as found in VirusTotal.

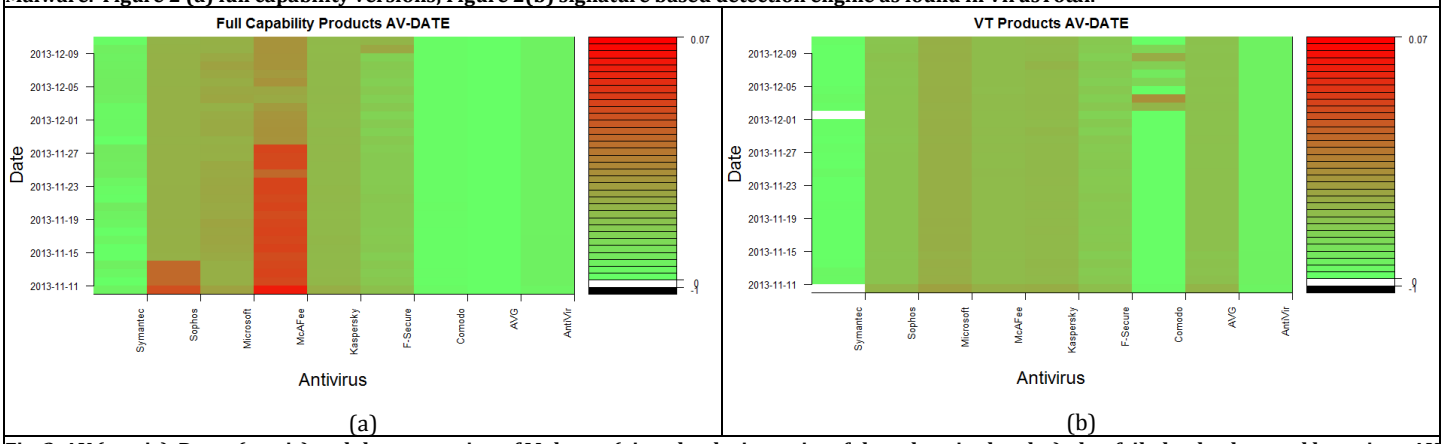

Fig. 3. AV (x-axis), Dates (y-axis) and the proportion of Malware (given by the intensity of the colour in the plot) that failed to be detected by a given AV on a given date. Figure 3 (a) full capability versions; Figure 3(b) signature based detection engine as found in VirusTotal.

malware to the two versions (the malware are sent to the VirusTotal versions on average two hours earlier than it is sent to the full capability versions, which should give a slight advantage to the full capability versions as they would work with a signature ruleset which is "fresher" by two hours).

Row A shows the count of undetected malware from both version, row B shows the count of undetected malware by full capability but detected by VirusTotal, and row C shows the count of the detected malware by Full Capability but not detected by VirusTotal, as we can see only AVG of VirusTotal could not detect 60 malware while the full capability version was able to detect them.

V. DISCUSSION

In this section we summarise the main observations we have made from our experiment and discuss the possible implications they may have on product selection and administration.

Observation: Out of nine vendors in our study, only two of them had a full capability version which had a detection rate that was better than their VirusTotal counterpart.
Implication: This suggests that for most of these products the free version they have in VirusTotal is perfectly suitable for malware detection and may even perform better compared with a full capability one. But we don't know yet to what extent these findings are condition to the setup under which the full capability products ran. We are currently building a more extensive infrastructure with full capability products to see whether we observe the same results.

Observation: The full capability versions of some of these AV vendors detected a malware in some cases more than three weeks after their VirusTotal version had detected the same malware.
Implication: This is a strange and counterintuitive observation. One would expect that the customers who have downloaded a paid version of a product would be served the signatures first. We speculate that the vendors are worried about false positives and want to first roll out a signature in VirusTotal. Only after they gain enough confidence that a file is indeed malicious, do they roll it out to the full capability versions.

VI. CONCLUSION AND FURTHER WORK

In this paper we reported results of an empirical analysis of the detection capabilities of 9 AV products when they were subjected to 3605 malware samples collected on an experimental network over a period of 31 days. We compared the detection capabilities of the version of the AV products that the vendors make available for free in VirusTotal versus the full capability products that they make available via their own website. We analysed the detection capabilities of these products over the entire collection period and also checked which versions of these vendors first detect the malware. The main conclusions we can draw from our analysis are:

- For most vendors in our study (seven out of nine) the VirusTotal version has a better detection rate than their full capability version. This would imply that investment in the full capability version of an AV product may not be worthwhile.
- Some of the full capability versions of the AV vendors in our study only detected some malware more than a week after the VirusTotal version of the same vendor has detected the same malware. This seems to imply that vendors for some malware are testing their detection signatures in their VirusTotal versions first before propagating them to the full capability versions, which may also explain the higher detection rates of the VirusTotal versions of some of these vendors.

The main limitations of our conclusions are:

- The malware samples are mainly windows portable executable files collected with Dionaea;
- All our analysis so far has been with malware samples, which means we cannot get any measurements on false positive rates;
- We have looked at 9 vendors over a one month period. A longer data collection time with more vendors may allow for stronger conclusions.

The provisions for further work are:

- Studying the detection capability with different categories of malicious files. Further studies are needed to check the detection capability for other types of files e.g. document files, media files etc., and using malware collected with more platforms than just Dionaea;
- Extend the experiment with different types of platform, e.g. Windows, Mac and Linux;
- Studying the detection capabilities with datasets that allow measurements of false positives in addition to false negative rates. Getting representative datasets that allow for measurements of false positive rates is difficult, but we will continue researching suitable mechanisms that would allow for these measurements to take place;
- Further work on the dynamic analysis of the malware, using more than just the malwr.com site;
- Discussions with the vendors to gain more insight (where possible) on their malware detections, signature writing conventions and propagation rules for these signatures to their different versions of products.

REFERENCES

[1] H. S. K. and J. C. H. Sukwong, O., "Commercial Antivirus Software Effectiveness: An Empirical Study," IEEE Computer, 2011. 44(3): p. 63-70.

[2] J. Oberheide, E. Cooke, and F. Jahanian, "CloudAV: N-version Antivirus in The Network Cloud," in *Proceedings of the 17th conference on Security symposium*, 2008, pp. 91–106.

[3] P. Bishop, R. Bloomfield, I. Gashi, and V. Stankovic, "Diversity for security: A study with off-the-shelf AntiVirus engines," in *Proc, ISSRE*, 2011, pp. 11–19.

[4] I. Gashi, V. Stankovic, C. Leita, and O. Thonnard, "An experimental study of diversity with off-the-shelf antivirus engines," in *Proceedings* of the 8th IEEE Int. Symp. on Network Computing and Applications (NCA), p. 4-11, 2009

[5] L. Spitzner, "Honeypots: Tracking Hackers," *Addison-Wesley Longman Publishing Co.,Inc*, 2002. .

[6] D. Ramsbrock, R. Berthier, and M. Cukier, "Profiling Attacker Behavior Following SSH Compromises," in *Proc. DSN'07*, 2007, pp. 119–124.

[7] F. Pouget, "Leurre.com: on the advantages of deploying a large scale distributed honeypot platform," 2005.

[8] C. Leita and M. Dacier, "SGNET: A worldwide deployable framework to support the analysis of malware threat models," in *Proc. EDCC-7*, 2008, pp. 99–109.

[9] C. Leita and M. Dacier, "SGNET: Implementation insights," in *NOMS 2008* - pp. 1075–1078.

[10] M. Vrable, J. Ma, J. Chen, et al "Scalability, fidelity, and containment in the potemkin virtual honeyfarm," in *Proc SOSP '05*, 2005, vol. 39, no. 5, p. 148.

[11] "Comparing Detection Capabilities of AntiVirus Products: Technical Report,"2016 available at: https://www.sugarsync.com/pf/D2156856_83687465_071572