Microsoft

Transparency report

Examining the AV-TEST November-December 2018 results

Prepared by

Windows Defender Research Team

Microsoft

Table of Contents

Microsoft

1 Report highlights

In AV-TEST's November-December 2018 testing cycle, Windows Defender Antivirus achieved a perfect score (6.0/6.0) in the Protection test. This is the fourth consecutive cycle that Windows Defender Antivirus achieved a perfect score.

Compared to last cycle, Windows Defender Antivirus' Usability test score dropped to 5.5/6.0, while the Performance test score increased to 5.5/6.0.

This report presents more details on test scores, with commentary for context and transparency.

1.1 Key takeaways

Below is a summary of this report:

Protection

Windows Defender Antivirus maintained an overall Protection score of 6.0/6.0, detecting 100% of 19,956 malware samples. With the latest results, Windows Defender Antivirus has achieved a perfect score in 5 out of 6 test cycles in 2018.

Usability (false positives)

Windows Defender Antivirus' Usability score dropped to 5.5/6.0 after misclassifying 5 out of 1,507,247 files tested.

Performance

Windows Defender Antivirus achieved an overall Performance score of 5.5/6.0, an improvement from its previous 5.0/6.0 score. It registered higher performance impact during software installation, a low-frequency action, but was better than the industry average in all other actions.

Testing methodology

Microsoft continues to observe misalignment between testing methodologies and the way threats occur in the real world. Microsoft is working with several testers to bridge the gap and drive more real-world testing.

2 Examining test results

2.1 Summary of overall scores

The table below summarizes overall test results for Windows Defender Antivirus in the November-December 2018 antivirus testing by AV-TEST:

	Protection	Usability	Performance
Overall scores for this cycle >>>	6.0/6.0 (±0)	5.5/6.0 (-0.5)	5.5/6.0 (+0.5)

Table 1. Windows Defender Antivirus' overall antivirus test results in the November-December 2018 AV-TEST Business User test. AV-TEST uses Protection, and Usability, and Performance test modules.

2.2 Understanding Protection scores

Below are details on the Protection test scores.

	November	December
"Real World" testing	100% (138/138)	100% (121/121)
"Prevalent malware" testing	100% (10,922/10,922)	100% (8,775/8,775)
Overall malware protection rate (all samples)	100% (19,956/19,956)	
Overall Protection score for this cycle >>>	**6.0/6.0 (±0)**	
Overall Protection ranking for this cycle >>	1st out of 16 (tied with 12 more)	

Table 2. Summary of Protection scores for the November-December 2018 Business User test

The diagrams below show Windows Defender Antivirus detection rates in “Prevalent Malware” and “Real World” testing over a one-year period. Windows Defender AV achieved 100% in 11 out of the 12 monthly “Prevalent malware” tests and 100% in 10 out of the 12 monthly “Real World” tests.

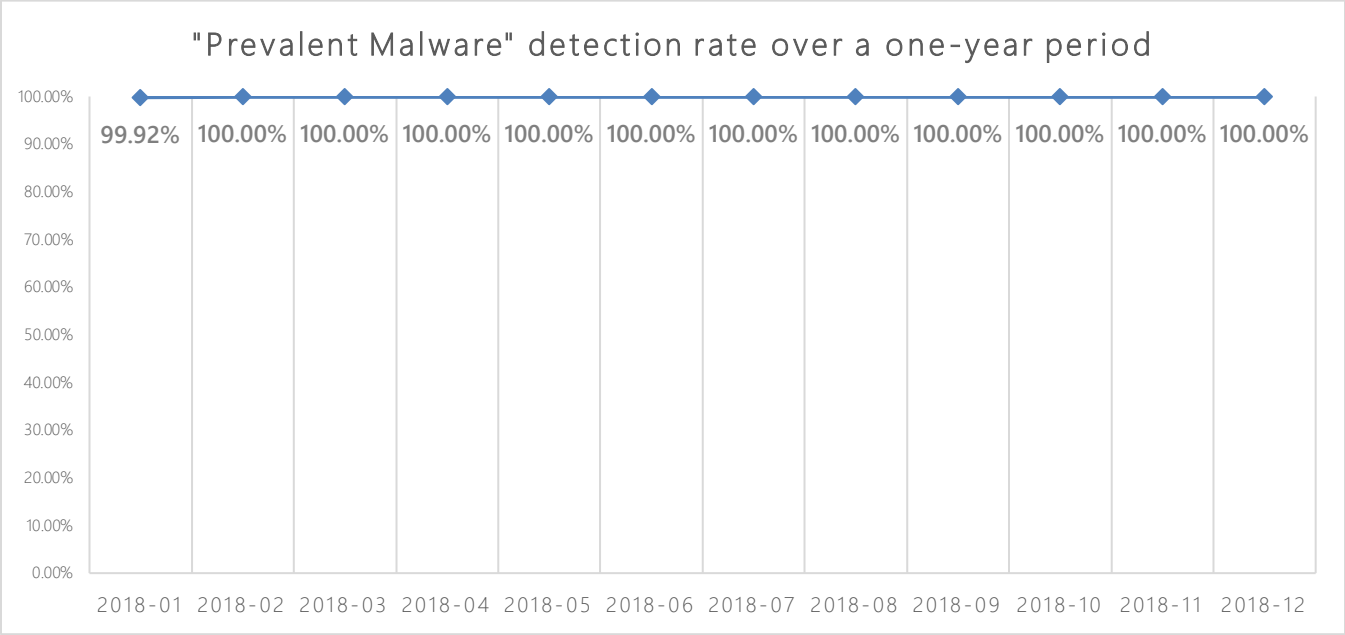

Figure 1. Windows Defender Antivirus detection rates in AV-TEST “Prevalent malware” tests over a one-year period

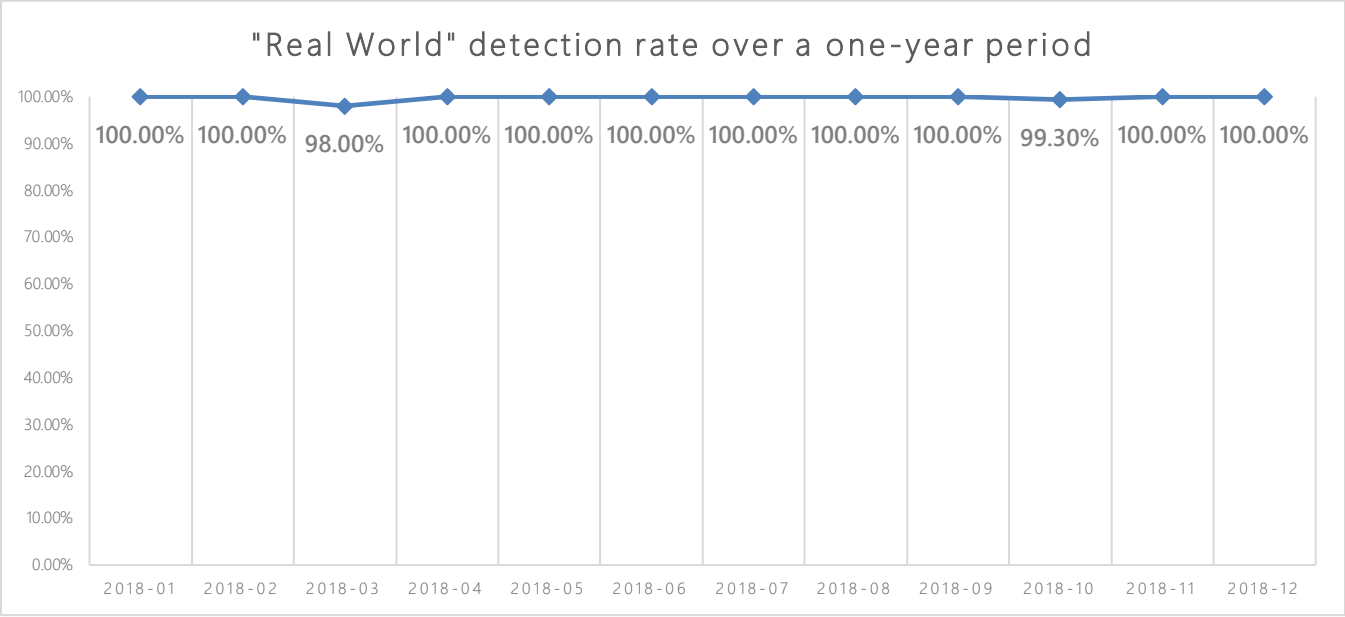

Figure 2. Windows Defender Antivirus detection rates in AV-TEST “Real World” tests over a one-year period

2.3 Understanding Usability scores

In Usability tests, AV-TEST includes clean file samples in the test population and checks whether antivirus products incorrectly classify them as malware (what is known as false positive, or FP). Below is a summary of results in the Usability test.

	November	December
Number of misclassified files	4 (out of 820,894 samples)	1 (out of 705,550 samples)
Overall Usability score for this cycle >>>	**5.5/6.0** (-0.5)	
Overall Usability ranking for this cycle >>>	14th out of 16 (tied with 1 more)	

Table 3. Summary of Usability test scores for the November-December 2018 Business User test

2.3.1 Analysis: What kinds of files were misclassified?

Below is a list of files that Windows Defender Antivirus misclassified in this test cycle. Based on our research and on file prevalence data, the misclassified samples are not common in enterprise environments.

Sample	File prevalence (30 days)	Description	Digitally signed? (Y/N)
Sample a	300	Dictionary application	N
Sample b	50	Benchmarking and diagnostic tool	N
Sample c	2	Document printing software	N
Sample d	10	App for customizing Windows install media	N
Sample e	1	App for customizing music mixes	N

Table 4. Files that Windows Defender antivirus incorrectly classified as malware

Microsoft encourages software vendors to take steps to raise the level of trust both by security vendors and users alike. These steps include signing software with certificates issued by reputable Certification Authorities.

2.3.2 The synthetic nature of usability tests

Misclassifications in a synthetic test are not necessarily indicative of false positives in real-world scenarios. This is true when the test methodology discounts contextual elements that Windows

Defender Antivirus uses for issuing a verdict. For example, when a file is tested, it is not downloaded from the vendor website. Both the original file name and the download site are contextual information that are removed in tests. We've seen many cases where a customer in the real world downloads a clean program from the vendor site without encountering any erroneous detection. However, when a tester gives the file a seemingly random name (e.g., its SHA-256 hash), removes the mark of the web, and doesn't download the file from the vendor website, some of our more aggressive machine learning models issue blocks that don't occur in the real world.

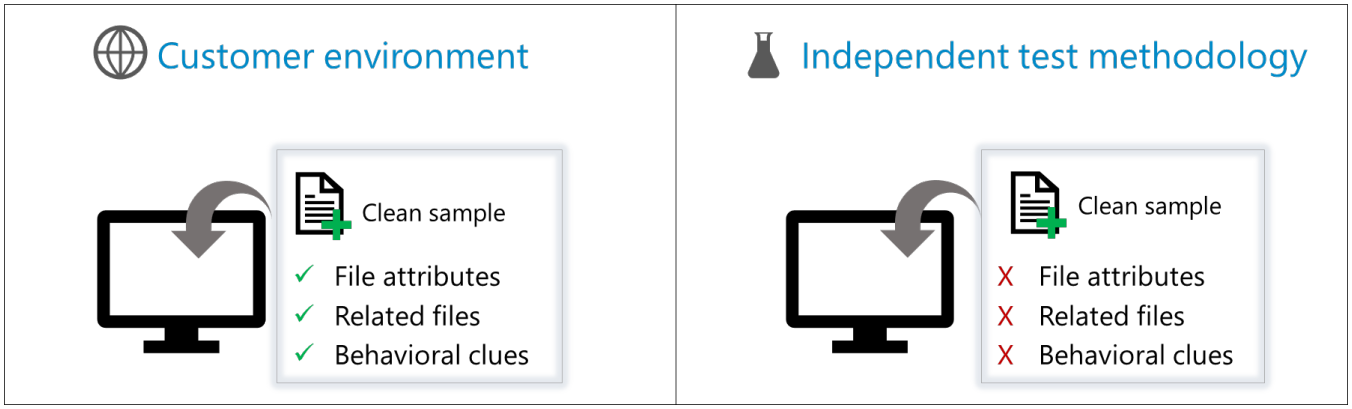

Figure 3. In some cases, samples are incorrectly classified (false positive) in the synthetic test environment but not on customer machines.

2.3.3 Criteria for evaluating files may vary across vendors and testers

The criteria for classification can vary between antivirus vendors and testers depending on their policies. Some files identified as clean by some vendors could be files that Windows Defender Antivirus identifies as potentially unwanted application (PUA) and thus would be blocked. Microsoft's policy aims to protect customers against malicious software while minimizing the restrictions on developers. The diagram below demonstrates the high-level evaluation criteria Microsoft uses for classifying samples:

- **Malicious software**: Performs malicious actions on a computer.
- **Unwanted software**: Exhibits the behavior of adware, browser modifier, misleading, monitoring tool, or software bundler
- **Potentially unwanted application (PUA)**: Exhibits behaviors that degrade the Windows experience
- **Clean**: We trust that the file is not malicious, is not inappropriate for an enterprise environment, and does not degrade the Windows experience

Figure 4. Microsoft's high-level sample classification criteria

2.3.4 We took notice: How the Windows Defender Antivirus team dealt with FPs

Our research team analyzed the samples that Windows Defender Antivirus misclassified and assigned proper determination. The team also analyzed the root causes for these misclassifications and worked with different threat research teams to enhance detection accuracy.

Below are some examples of detection improvements that research teams have made or are making in response to FPs made in the latest test:

- Retraining machine learning models for more accurate classifications
- Improving pre-release test processes to detect and correct any FPs ahead of time
- Improving monitoring to detect FP behavior using cloud protection

2.4 Understanding Performance scores

Performance tests measure the effect of certain user actions, which are executed as part of the test, on system speed. The table below summarizes Performance test results.

	November-December
Performance test score for this cycle	5.5/6.0 (+0.5)
Performance ranking for this cycle	2nd out of 16 (tied with 5 more vendors)

Table 5. Summary of Performance test scores for the November-December 2018 Business User test

The table below presents Windows Defender Antivirus' performance test results compared to industry averages. Performance is measured by the average impact of the product on computer speed; therefore, a smaller number is favorable. Green boxes indicate areas where Windows Defender Antivirus performed better than the industry average; red boxes indicate performance lower than the industry average.

Action	Standard PC	Industry average	High End PC	Industry average
Launching popular websites	9%	20%	6%	19%
Downloading frequently used applications*	0%	1%	0%	1%
Launching standard software applications	8%	11%	8%	9%
Installation of frequently used applications	51%	33%	34%	31%
Copying of files (locally and in a network)	1%	3%	1%	5%

Table 1. Average impact of the product on computer speed in daily usage

*The description for these operations is given by AV-TEST and might not be aligned with what Microsoft's data indicates as realistic.

2.4.1 Areas that matter the most to customers

Based on results presented in Table 6, Windows Defender Antivirus performed better than the industry average in all areas except in the area that AV-TEST labels as "*Installation of frequently-used applications*". There are several factors to consider for driving the right conclusion out of these test results:

- **Consider the frequency of the action**
  Most users in enterprise environments are information workers whose common user activities include:
  - Browsing the web
  - Using email clients
  - Processing documents
  - Accessing network resources

  Users spend substantially less time installing new applications compared to the activities listed above. This is true for all user segments, but especially for enterprises, where software installation is usually governed by usage policies. Windows Defender Antivirus is optimized for delivering high levels of performance during high-frequency actions. For instance, Windows Defender Antivirus performed lower than the industry average in *Installation of frequently used applications*, a low-frequency action, but it performed better than average in all other actions.

Performance is a priority area for the Windows Defender Antivirus team, and we're working to improve it even further.

- **Consider the level of risk**
  Windows Defender Antivirus is designed to perform thorough scanning during the software installation process. This could have a performance cost. One reason for this is that software installation is a relatively complex operation that touches different areas of the operating system. Thorough inspection is necessary to reduce the risk of introducing malicious software on the system.

- **What impactful areas are not being tested?**
  There are several areas that are not being tested for performance by AV-TEST that are critical to user experience. Examples include:

  - Shutdown and startup
  - Universal Windows app launch
  - Battery consumption

2.4.2 Improvements made in this cycle

The Windows Defender Antivirus team investigates performance logs generated in third-party tests and looks for opportunities to improve performance. Based on on the team's findings, we recently identified a technique to improve how Windows Defender Antivirus handles scanning of archive files. The engineering team is working to apply this improvement and will monitor the extent to which this technique could improve performance.