

# Transparency report

## Examining the AV-TEST July-August results

*Prepared by*

Windows Defender Research Team

© 2018 Microsoft Corporation. All rights reserved. This document is provided "as-is." Information and views expressed in this document, including URL and other Internet Web site references, may change without notice. You bear the risk of using it. This document does not provide you with any legal rights to any intellectual property in any Microsoft product. You may copy and use this document for your internal, reference purposes.

The descriptions of other companies' products in this document, if any, are provided solely as a convenience to aid understanding and should not be considered authoritative or an endorsement by Microsoft. For authoritative descriptions of any non-Microsoft products described herein, please consult the products' respective manufacturers.

Any use or distribution of these materials without the express authorization of Microsoft is strictly prohibited.

Microsoft and Windows are either registered trademarks or trademarks of Microsoft in the United States and/or other countries.

# Table of Contents

1	Introduction.....	2
1.1	Report highlight.....	2
1.2	Key takeaways.....	2
2	Examining test results .....	3
2.1	Summary of overall scores.....	3
2.2	Understanding Protection scores .....	3
2.3	Understanding Usability Scores.....	5
2.3.1	Analysis: What kinds of files were misclassified? .....	5
2.3.2	The synthetic nature of usability tests .....	5
2.3.3	Criteria for evaluating files may vary across vendors and testers.....	6
2.3.4	We took notice: How the Windows Defender Antivirus team dealt with FPs.....	7
2.4	Understanding Performance scores.....	7
2.4.1	Areas that matter the most to customers .....	8

# 1 Introduction

In AV-TEST's [July-August 2018](#) testing cycle, [Windows Defender Antivirus](#) achieved perfect scores (6.0/6.0) in the Protection, Usability, and Performance test modules. This report presents more details on test scores, with commentary for context and transparency.

## 1.1 Report highlight

Windows Defender Antivirus achieved perfect scores in all three categories: Protection, Usability, and Performance.

## 1.2 Key takeaways

Below is a summary of this report:

1

### Protection

Windows Defender Antivirus maintained an overall Protection score of 6.0/6.0, detecting 100% of 20,022 malware samples. With the latest results, Windows Defender Antivirus has achieved 100% on 14 of the 16 antivirus tests (combined "Real World" and "Prevalent malware") in 2018.

2

### Usability (false positives)

Windows Defender Antivirus also maintained a Usability score of 6.0/6.0 after misclassifying only 2 files out of 1,373,427 files tested.

3

### Performance

Windows Defender Antivirus achieved an overall Performance score of 6.0/6.0, an improvement from its previous 5.5/6.0 score. The results showed slightly higher performance impact in low frequency actions (e.g., software installation).

4

### Testing methodology

Microsoft continues to observe areas to improve alignment between testing methodologies and the way threats occur in the real world. Microsoft is working with a number of testers to bridge the gap and drive true 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 July-August 2018 antivirus testing by AV-TEST:

	Protection	Usability	Performance
Overall score for this cycle >>>	6.0/6.0 ( $\pm 0$ )	6.0/6.0 ( $\pm 0$ )	6.0/6.0 (+0.5)

Table 1. Windows Defender Antivirus' overall antivirus test results in the [July-August 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:

	July	August
"Real World" testing	100% (118/118)	100% (157/157)
"Prevalent malware" testing	100% (6,580/6,580)	100% (13,167/13,167)
Overall malware protection rate (all samples)	100% (20,022/ 20,022)	
Overall Protection score for this cycle >>>	<b>6.0/6.0 (<math>\pm 0</math>)</b>	
Overall Protection ranking for this cycle >>	1 <sup>st</sup> out of 18 (tied with 16 more)	

Table 2. Summary of [Protection](#) scores for the July-August 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:

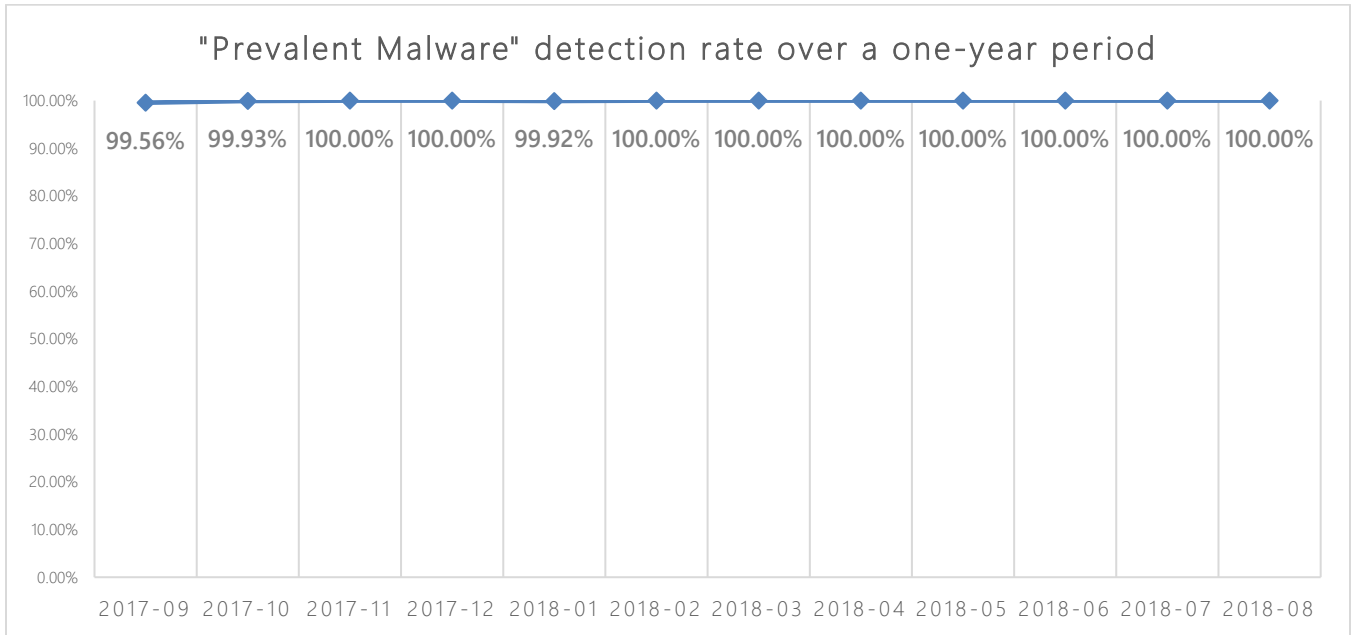


Figure 1. Windows Defender Antivirus detection rates in AV-TEST "Prevalent malware" tests over a one-year period

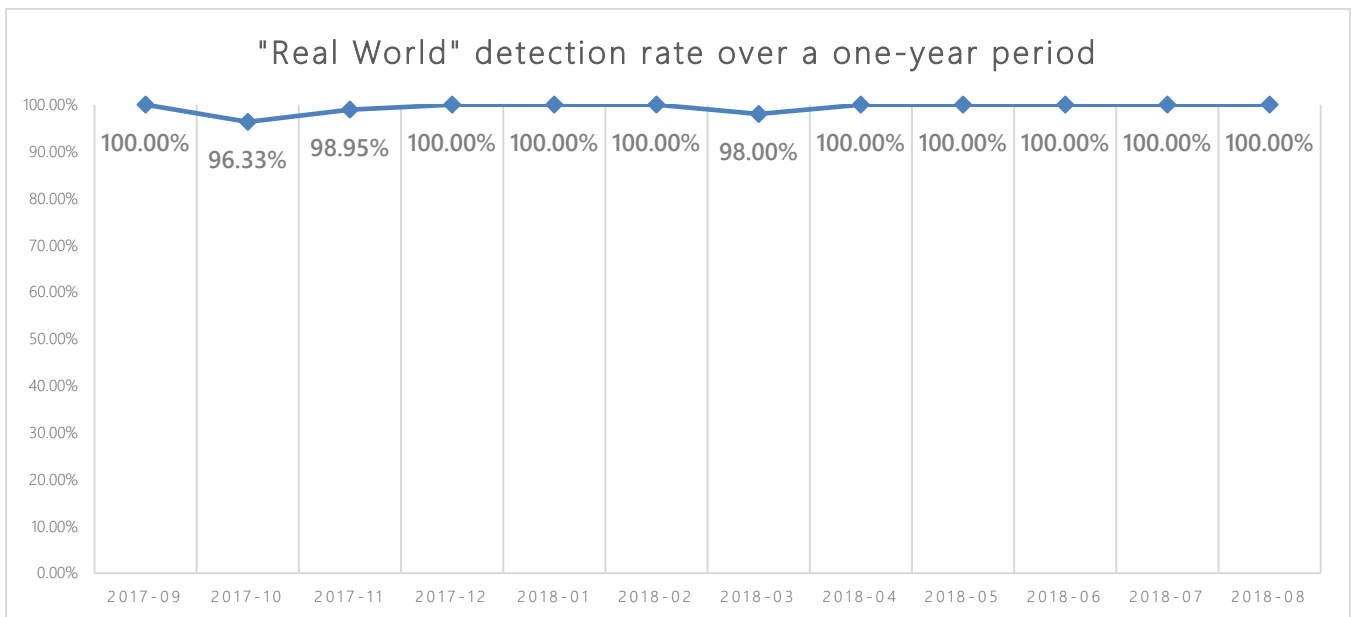


Figure 2. Windows Defender Antivirus detection rates in AV-TEST "Real World" tests over a one-year period

## 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:

	May	June
Number of misclassified files	0 (out of 796,381 samples)	2 (out of 577,046 samples)
Overall Usability score for this cycle >>>	<b>6.0/6.0 (±0)</b>	
Overall Usability ranking for this cycle >>>	1 <sup>st</sup> out of 18 (tied with 8 more)	

Table 3. Summary of [Usability test](#) scores for the July-August 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 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	Movie catalog software	N
Sample b	2	Duplicate file finder software	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.

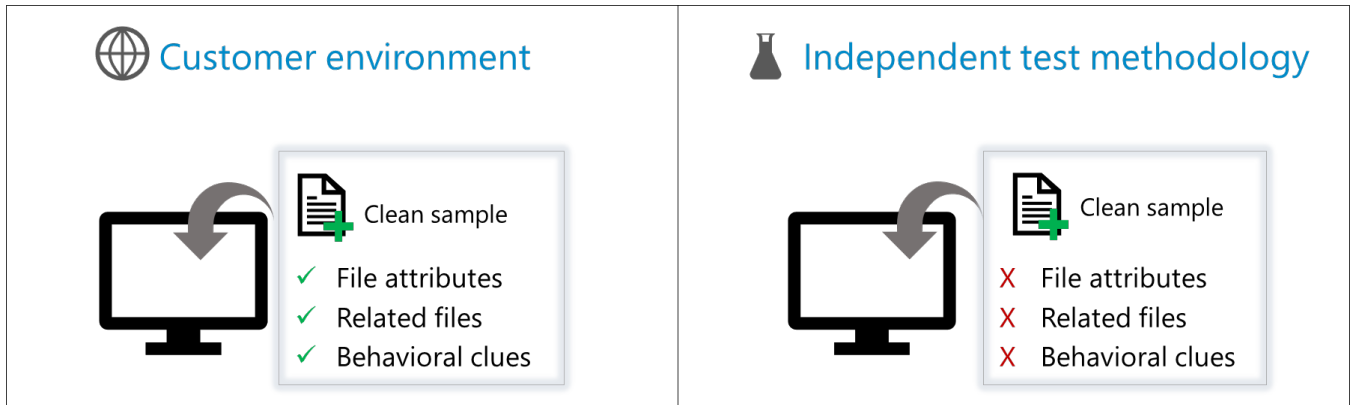


Figure 3. In some cases, samples are incorrectly classified (false positive) in the synthetic test environment but not on customer machines.

### 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

For each false positive, a specialized team analyzes the sample and assigns a proper determination. The team also analyzes the root cause for the misclassification and often works with the different threat research teams to enhance detection accuracy.

Below are some examples of the detection improvements that the research teams have made or are making in response to FPs made in the most recent AV-TEST test cycle:

- Retraining machine learning models for more accurate detections
- Improving pre-release test processes to detect any FPs ahead of time and correct them
- 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 in the July-August cycle:

	July-August
Performance test score for this cycle	6.0/6.0 (+0.5)
Performance ranking for this cycle	1 <sup>st</sup> out of 17 (tied with 8 more vendors)

Table 5. [Performance test](#) results for Windows Defender Antivirus for the July-August cycle

The table below presents the details of 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 lower than the industry average.

Action	Standard PC	Industry average	High End PC	Industry average
Launching popular websites	4%	18%	5%	13%
Downloading frequently-used applications*	2%	1%	0%	1%
Launching standard software applications	8%	11%	13%	12%
Installation of frequently-used applications	30%	26%	25%	23%
Copying of files (locally and in a network)	2%	6%	3%	11%

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 several areas. For the areas where it slightly underperformed, the largest gap between Windows Defender Antivirus' performance and the industry average (4%) is 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 example, *Installation of frequently used applications* (a low-frequency action) is the only area where Windows Defender

Antivirus scored lower than the industry average. 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