Microsoft

# Transparency report

## Examining the AV-TEST January-February 2018 Results

*Prepared by*

Windows Defender Research team

# Table of contents

# 1 Introduction

This report presents Windows Defender Antivirus (Windows Defender AV) test scores in the recent AV-TEST testing cycle (January-February 2018).

## 1.1 Report highlights

While maintaining a **consistent perfect score** in the Protection category, Windows Defender AV showed improvement in both Performance and Usability categories.

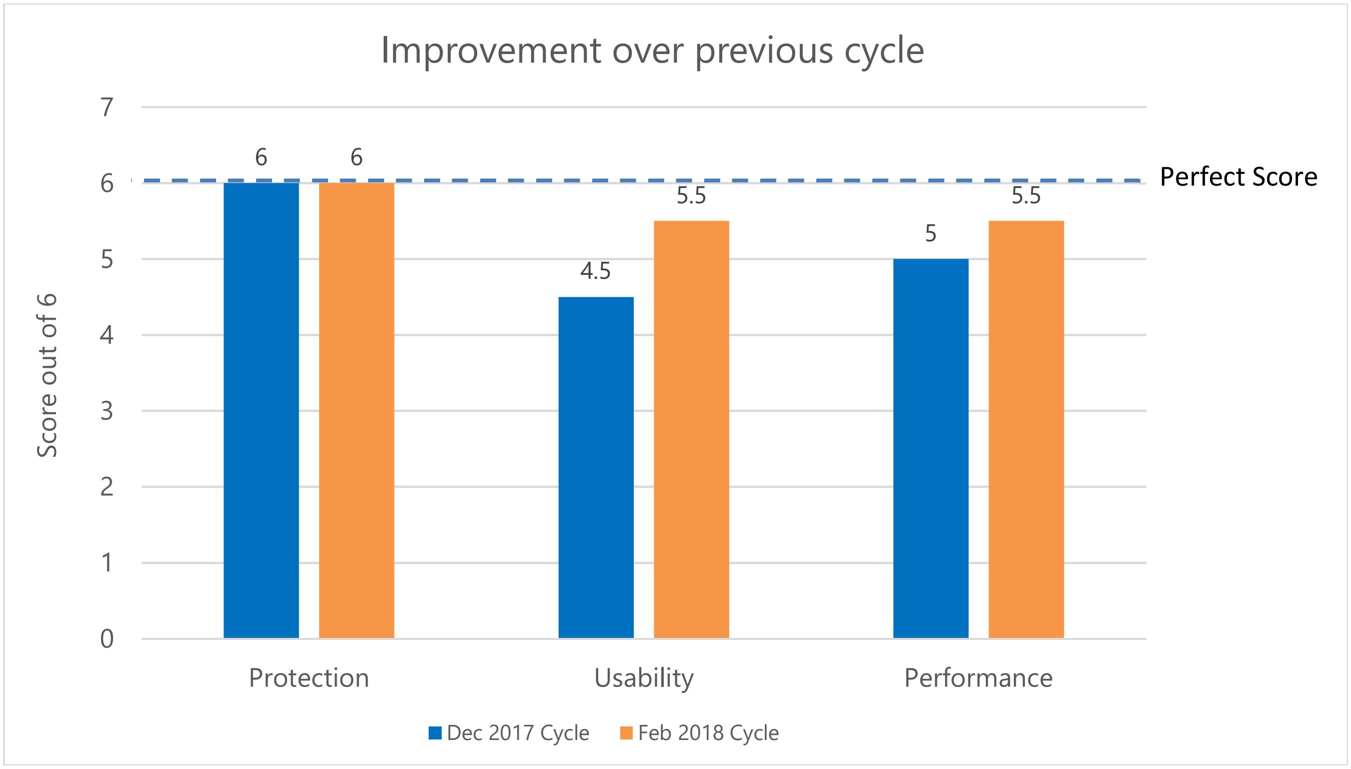


Figure 1. Improvement over previous cycle

## 1.2 Key takeaways

Below is a list of key takeaways from this report:

1. **Protection:** Windows Defender AV achieved a perfect score in Protection, maintaining a very competent trend in this category. Learn More

2. **Usability (false positives):** Usability improved in this test cycle. Windows Defender AV achieved a Usability score of 5.5/6.0. Per our telemetry, samples that Windows Defender AV incorrectly classified (false positive) had very low prevalence and are not commonly used in business context. Learn More

3. **Performance:** Windows Defender AV improved this cycle, achieving a 5.5/6.0 Performance score and outperforming the industry in almost all areas. These results reflect the investments we put in optimizing Windows Defender AV performance for high-frequency actions (e.g., application run). Learn More

# 2 Dissecting test results

## 2.1 Protection scores: Perfection maintained!

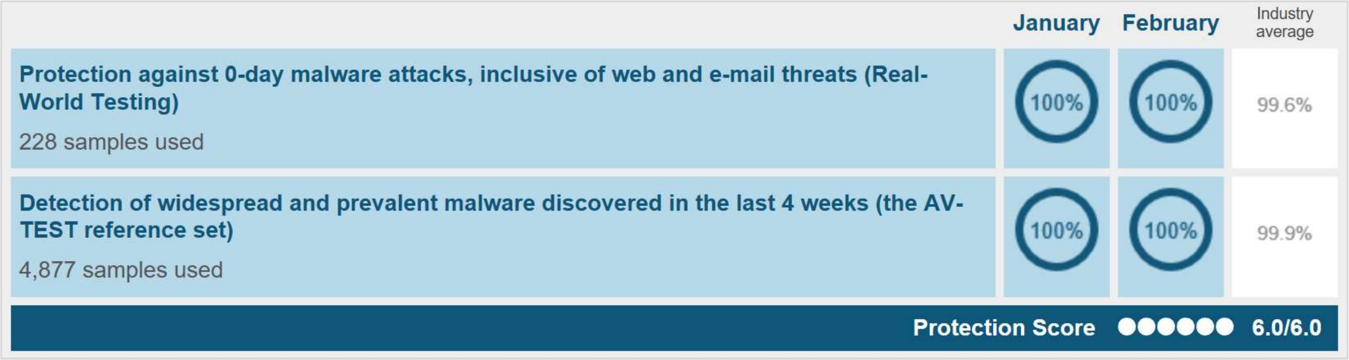
Below is a summary of Protection scores for the most recent AV-TEST Business User test:

| | January | February |
|---|---|---|
| "Real World" testing | 100% (125/125) | 100% (103/103) |
| "Prevalent" malware testing | 99.93% (2,710/2,712) | 100% (2,165/ 2,165) |
| Overall Protection score for this cycle >>> | **6.0/6.0** (±0)⬌ | |
| Overall Protection ranking for this cycle >> | 1st out of 16 (tied with 11 more vendors) | |

Table 1. Summary of Protection scores for the January-February 2018 Business User test

**Note:**

In the Home User test, Windows Defender AV scored 100% in both "Real World" and "Prevalent" testing in both January and February



| | January | February | Industry average |
|---|---|---|---|
| **Protection against 0-day malware attacks, inclusive of web and e-mail threats (Real-World Testing)**<br>228 samples used | 100% | 100% | 99.6% |
| **Detection of widespread and prevalent malware discovered in the last 4 weeks (the AV-TEST reference set)**<br>4,877 samples used | 100% | 100% | 99.9% |
| | | **Protection Score** ●●●●●● | **6.0/6.0** |

Figure 2. Perfect Protection scores in the Home User test (from www.av-test.org)

### 2.1.1 Two missed samples, two opportunities for improvement

The Windows Defender Research team takes missed samples as an opportunity to improve detection capabilities. For each missed sample, a team of researchers analyzes and assigns a proper label to the sample to make sure it is detected. In addition, the team also analyzes the root cause for the miss and drives long-term detection improvements.

Windows Defender AV missed two samples in this cycle. Below is the analysis of the miss and the improvements that were introduced as a result:

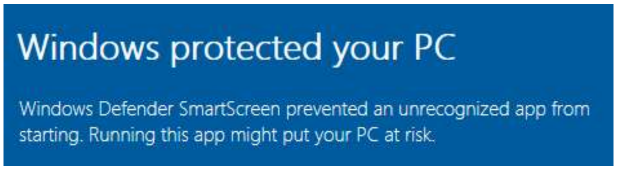
| Missed sample | Miss reason | Actions taken |
|---|---|---|
| Sample 1 | • Incorrect classification | • Retrained machine learning classifiers |
| Sample 2 | • No classification | • Retrained machine learning classifiers |

Table 2. Improvements made to Windows Defender AV in response to this cycle's results

### 2.1.2 True real-world Testing: Testing against the Windows Defender ATP stack

The Windows Defender AV team tested the two missed samples against the Windows Defender Advanced Threat Protection (Windows Defender ATP) stack to assess the samples' ability to infect machines in real-world enterprise environments. This expands on the testing practice that isolates AV from the rest of the environment, which eliminates the synergy between stack components and creates synthetic conditions. As expected, the malware samples were blocked and detected by several stack components, as follows:

**Sample 1**:

| Windows Defender ATP component | Outcome |
|---|---|
| Windows Defender SmartScreen | "Unknown program" prompt displayed upon download attempt<br><br>Windows protected your PC<br>Windows Defender SmartScreen prevented an unrecognized app from starting. Running this app might put your PC at risk. |
| Windows Defender Application Control | Blocked from running under the following modes:<br>- Intelligent Security Graph mode<br>- Whitelisting mode<br>- Managed Installer mode |
| Windows Defender Application Guard | Windows Defender Application Guard blocks this threat from being downloaded and run from the web |

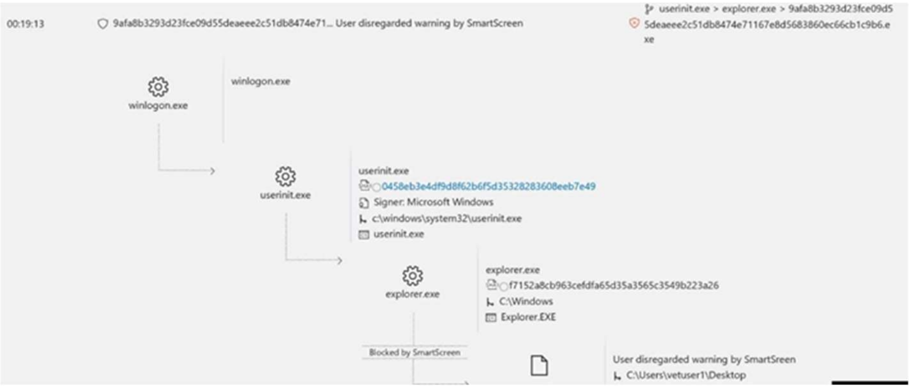
| | |
|---|---|
| Windows Defender ATP Endpoint Detection and Response (EDR) | Alert generated: “User disregarded warning by SmartScreen”<br><br>00:19:13 9afa8b3293d23fce09d55deaeee2c51db8474e71... User disregarded warning by SmartScreen<br>userinit.exe > explorer.exe > 9afa8b3293d23fce09d55deaeee2c51db8474e71167e8d5683860ec66cb1c9b6.exe<br>winlogon.exe<br>winlogon.exe<br>userinit.exe<br>userinit.exe<br>0458eb3e4df9d8f62b6f5d35328283608eeb7e49<br>Signer: Microsoft Windows<br>c:\windows\system32\userinit.exe<br>userinit.exe<br>explorer.exe<br>explorer.exe<br>f7152a8cb963cefdfa65d35a3565c3549b223a26<br>C:\Windows<br>Explorer.EXE<br>Blocked by SmartScreen<br>User disregarded warning by SmartSreen<br>C:\Users\vetuser1\Desktop |

Table 3. Running sample 1 against the Windows Defender ATP stack

**Sample 2**:

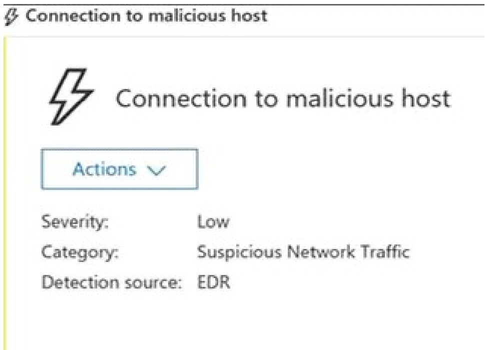
| Windows Defender ATP component | Outcome |
|---|---|
| Windows Defender Application Control | Blocked from running under the following modes:<br>- Whitelisting mode<br>- Managed Installer mode |
| Windows Defender Application Guard | Windows Defender Application Guard blocks this threat from being downloaded and run from the web |
| Windows Defender ATP Endpoint Detection and Response (EDR) | Alert generated: “Connection to malicious host”<br><br>Connection to malicious host<br>Connection to malicious host<br> <br>Severity: Low<br>Category: Suspicious Network Traffic<br>Detection source: EDR |

Table 4. Running sample 2 against the Windows Defender ATP stack

## 2.2 Improvement in Usability scores

In Usability tests, AV-TEST includes clean file samples in the test population and checks whether AV products falsely classify them as malware (what is known as a False Positive, or FP).

Below is a summary of Windows Defender AV results in the Usability test

| | January | February |
|---|---|---|
| Number of FPs | 2 | 4 |
| Overall score for this cycle >>> | **5.5/6.0** (+1.5)⬆ | |
| Ranking | 8th out of 16 vendors (tied with 3 more vendors) | |

Table 5. Usability test summary for Windows Defender AV for the Jan-Feb 2018 cycle in the Business User test

**Note**: Windows Defender AV scored 6.0/6.0 in the Home User test.

### 2.2.1 Analysis: What kind of files did Windows Defender AV misclassify?

Below is a sample list of files that Windows Defender AV incorrectly classified (false positive) on in this test cycle. Based on our research, most of the samples are not typical to enterprise environments (see File prevalence column).

| Sample | Description | File prevalence (30 days) |
|---|---|---|
| *Sample a* | SMS dispatch tool | 0 |
| *Sample b* | Media player | 19 |
| *Sample c* | Audio mixer | 2 |

Table 6. Files that Windows Defender AV incorrectly classified as malware for the January-February 2018 cycle

Microsoft encourages independent software vendors to sign their software with certificates issued by reputable Certification Authorities. This will raise the level of trust both by security vendors and users alike.

### 2.2.2 Synthetic testing conditions

Synthetic testing can be tricky. False positives in a synthetic test are not necessarily indicative of false positives in real-world scenarios. One potential way this can happen is when the test methodology discounts contextual elements that an AV product uses for issuing a verdict. For example, when a file is tested, it is not downloaded from the vendor website. Both the original name and the download site are contextual information that are removed in synthetic tests. We've seen many cases where a customer downloaded 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., SHA-256), removes the Mark of the Web, and doesn't download the file from the vendor website, some of our more aggressive machine learning models issued blocks that don't occur in the real world .

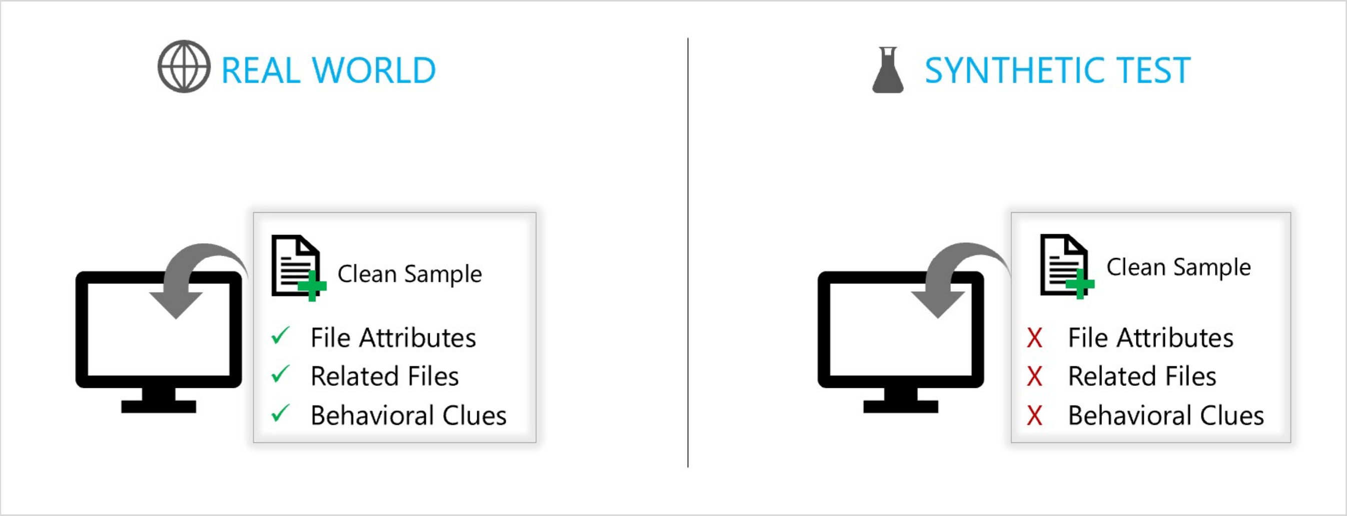


Figure 3. In some cases, Windows Defender AV incorrectly classified samples (false positive) in the synthetic test environment but not on customer machines

## 2.3 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 Windows Defender AV's results in the January-February cycle:

| | January-February |
|---|---|
| Overall Performance test score | **5.5/6.0** (+0.5) ↑ |
| Windows Defender AV ranking | 7th/15 (tied with four more vendors) |

Table 7. Summary of Windows Defender AV scores in the Performance tests for the Jan-Feb 2018 cycle

The table below shows 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 AV performed better than or the same as the industry average; red boxes indicate where Windows Defender AV performed lower than the industry average.

| Action | Standard PC | Industry Average | High End PC | Industry Average |
|---|---|---|---|---|
| Launching popular websites | 4% | 16% | 8% | 14% |
| Downloading frequently used applications* | 0% | 3% | 0% | 2% |
| Launching standard software applications | 13% | 13% | 15% | 15% |
| Installation of frequently used applications | 53% | 27% | 39% | 29% |
| Copying of files (locally and in a network) | 1% | 3% | 1% | 6% |

Table 8. Average impact of the product on computer speed in daily usage for the Jan-Feb 2018 cycle

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

### 2.3.1 Key points to consider about Performance results

Based on the results presented in Table 8, Windows Defender AV outperforms the industry average in many areas. The only area where Windows Defender AV performance is below the industry average is in performance during 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 e-mail 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 it is especially true for enterprise users where software installation is usually governed by usage policies. Windows Defender AV's performance is optimized for delivering high levels of performance in high frequency actions for better overall user experience. This is evident in the data presented in Table 8 where *application installation* (a low-frequency action) is the only area where Windows Defender AV scored substantially lower than the industry average, while scoring higher than industry average in other areas.

- **Consider the level of risk**
  Windows Defender AV 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 address 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 the user experience. Examples include:

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