Microsoft

Transparency report

Examining industry test results, May 2019

Prepared by

Microsoft Defender ATP Research Team

Table of contents

1 Summary of latest industry test results

This report provides a review of the latest independent industry tests results for Windows Defender Antivirus, the next-generation protection component of Microsoft's unified endpoint protection platform, Microsoft Defender Advanced Threat Protection (Microsoft Defender ATP).

Over the last few years, Microsoft Defender ATP has continuously improved its performance in industry tests. Today, it consistently achieves top scores in these tests, showing the investments and innovations we've made and continue to make in our security technologies.

While current antivirus tests don't necessarily reflect how attacks operate and how solutions are deployed in real customer environments, test results can influence important business decisions. We're actively working with several industry testers to evolve security testing. Meanwhile, we're publishing this report to provide more details, insights, and context on results. We'd like to be transparent to our customers and to the industry about our wins as well as improvement plans resulting from these tests.

1.1 AV-TEST: Perfect Protection score (January-February 2019)

In AV-TEST's January-February 2019 testing cycle, Windows Defender Antivirus achieved a perfect score (**6.0/6.0**) in the Protection test. This is the fifth consecutive cycle that Windows Defender Antivirus achieved a perfect score.

Both Usability and Performance test scores remain at **5.5/6.0**, the same as last cycle. This report provides more details on AV-TEST scores, with commentary for context and transparency. Learn More >>

1.2 SE Labs: AAA Award (October-December 2018)

In SE Labs' Enterprise Endpoint Protection test in October-December 2018, Windows Defender Antivirus won the AAA Award. It also got the AAA Award in the previous two test periods.

Out of 11 solutions tested, Windows Defender Antivirus was 1 of only 2 vendors that achieved **100%** rating for both Protection Accuracy and Legitimate Accuracy, combining to a Total Accuracy rating of **100%**.

SE Labs determined that Windows Defender Antivirus was one of the most effective, citing "Microsoft achieved extremely good results due to a combination of their ability to block malicious URLs, handle exploits and correctly classify legitimate applications and websites." Learn More >>

1.3 AV-Comparatives: Approved Business Product (December 2018)

In April 2019, AV-Comparatives released results for the Business Security Test 2019 (March-April 2019). Windows Defender Antivirus achieved a protection rate of **99.7%** in the Real-world protection test (March-April) and **99.5%** in the Malware protection test (March).

Windows Defender Antivirus was certified Approved Business Product in December 2018 after garnering strong scores in the Business Security Test 2018 (August-November 2019). AV-Comparatives gives the Approved Business Product award twice a year. Learn More >>

2 Examining the AV-TEST results

2.1 Summary of overall AV-TEST scores

The table below summarizes the overall test results for Windows Defender Antivirus in the January-February 2019 antivirus testing by AV-TEST:

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

Table 1. Windows Defender Antivirus' overall antivirus test results in the January-February 2019 AV-TEST Business User test. AV-TEST uses Protection, and Usability, and Performance test modules.

2.2 Understanding Protection test scores

Below are more details on the Protection test scores.

	January	February
Real World testing	100% (144/144)	100% (165/165)
Prevalent malware testing	99.9% (11,216/11,222)	100% (2,445/2,446)
Overall malware protection rate (all samples)	100% (13,970/13,977)	
Overall Protection score for this cycle >>>	**6.0/6.0 (±0)**	
Overall Protection ranking for this cycle >>	1st out of 15 (tied with 10 more)	

Table 2. Summary of Protection scores for the January-February 2019 Business User test – not a miss to us in the February Prevalent testing

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

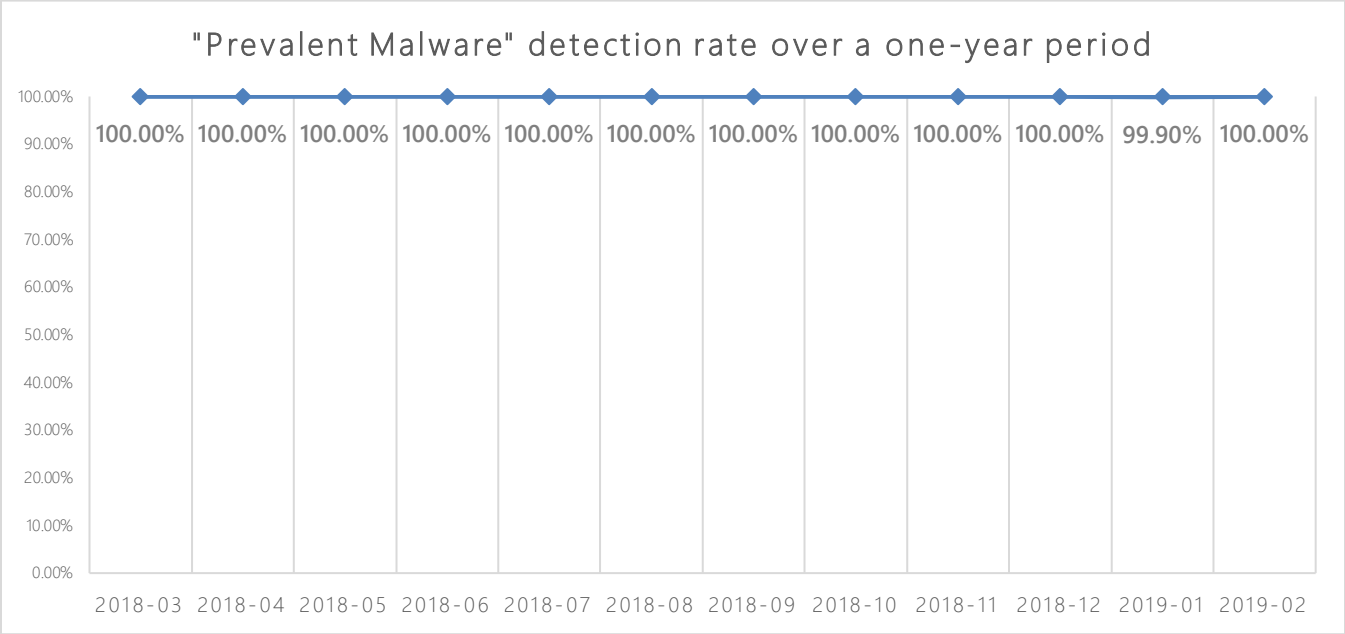

Figure 1. Windows Defender Antivirus detection rates in AV-TEST Prevalent malware tests over a one-year period

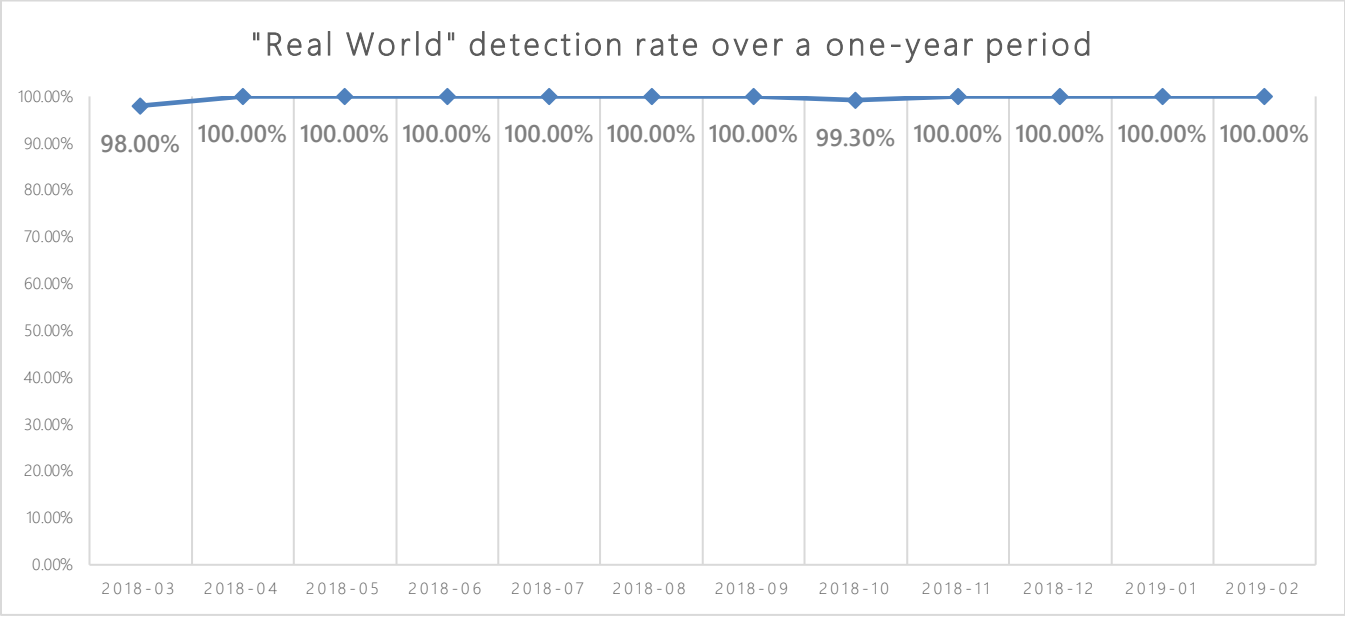

Figure 2. Windows Defender Antivirus detection rates in AV-TEST Real World tests over a one-year period

2.2.1 Missed samples are opportunities for improvement

Windows Defender Antivirus missed 7 out of 13,977 tested samples in Protection tests. The Windows Defender ATP team takes missed samples as an opportunity to improve detection capabilities. A team of researchers analyzes and assigns a proper label to the samples to make sure they are detected in the future. In addition, the team also analyzes the root cause for the misses and drives long-term detection improvements.

Below is a summary of the improvements that were introduced as a result of the misses:

Missed sample	Improvement made
Sample 1	Improved cloud-based protections for this category of threat
Sample 2	Improved internal workflow
Sample 3	Fixed a bug in cloud protecting service
Sample 4	Improved cloud-based protections for this category of threat
Sample 5	Improved cloud-based protections for this category of threat
Sample 6	Improved cloud-based protections for this category of threat
Sample 7	This sample is determined as clean per our classification policy; this is one example where tester policy and vendor policy don't align. Refer to section 2.3.3 for a broader commentary on this topic

Table 3. Improvements made to Windows Defender Antivirus in response to this cycle's results

2.3 Understanding Usability test 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.

	January	February
Number of misclassified files	5 (out of 805,754 samples)	4 (out of 800,169 samples)
Overall Usability score for this cycle >>>	**5.5/6.0** (±0)	
Overall Usability ranking for this cycle >>>	14th out of 15	

Table 4. Summary of Usability test scores for the January-February 2019 Business User test

2.3.1 Analysis: What kind 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	Global File prevalence (30 days)	Description	Digitally signed? (Y/N)
Sample a	0	Anti-malware app component	Y (signed with a revoked cert)
Sample b	10	Finance app	N
Sample c	248	Deployment tool	N
Sample d	118	Virus cleaner utility	N
Sample e	10	Custom user tool	N
Sample f	201	Media player app	Y (signed with a revoked cert)
Sample g	400	Media player app	Y (signed with a revoked cert)
Sample h	150	Text editor app	N
Sample i	2	Game app	N

Table 5. 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.

2.3.2 The synthetic nature of usability tests

Misclassifications (false positives) in a synthetic test are not necessarily indicative of false positives in real-world scenarios. This is because test methodologies often discount 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 (for example, 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.

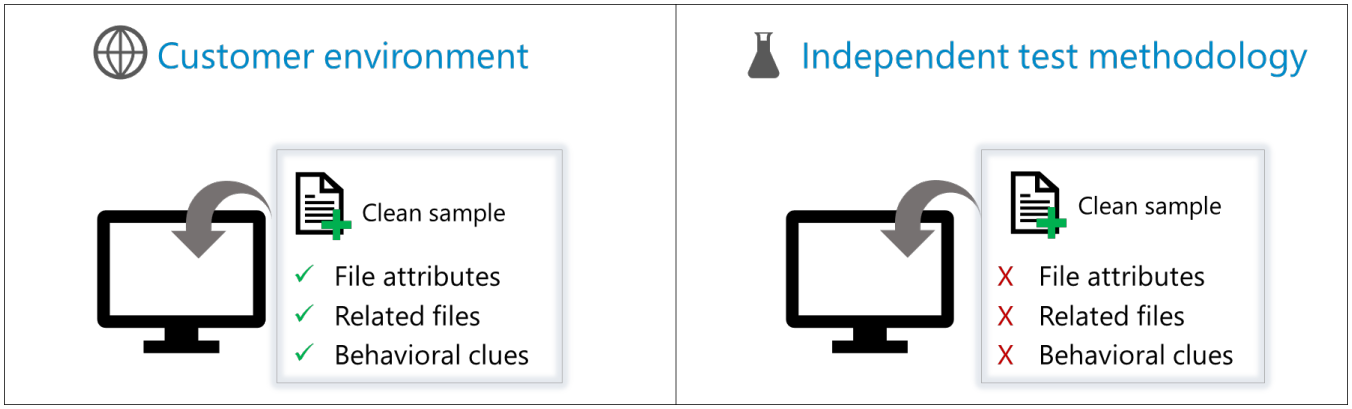

Figure 3. In some cases, samples are incorrectly classified (false positive) in the synthetic test environment but not on customer machines.

2.3.3 Files signed with revoked certificates

Three out of the 9 samples that were misclassified by Windows Defender Antivirus were signed with a revoked digital certificate. Revoked digital certificates should not be trusted, as they may indicate certificate or certificate authority (CA) compromise.

We encourage software vendors to rely on trustworthy certificate authorities for signing software and to inform software users about compromised certificates and CAs that they may have used for signing their software.

Despite the 3 files being classified clean by AV-TEST, we advise customers to be very cautious about using software signed with a revoked certificate.

2.3.4 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

2.3.5 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 to enhance detection accuracy.

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

- Refined cloud malware classification thresholds to improve the balance between protection and usability
- Added machine learning models that classify clean files; adjusted existing model weights to learn from previous FPs
- Expanded sources of clean file reputation

2.4 Understanding Performance test 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.

	January-February
Performance test score for this cycle	5.5/6.0 (+0.0)
Performance ranking for this cycle	6th out of 15 (tied with 7 more vendors)

Table 6. Summary of Performance test scores for the January-February 2019 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	4%	*20%*	4%	*16%*
Downloading frequently used applications*	2%	*1%*	0%	*1%*
Launching standard software applications	10%	*12%*	11%	*10%*
Installation of frequently used applications	61%	*33%*	55%	*30%*
Copying of files (locally and in a network)	1%	*3%*	0%	*4%*

Table 7. 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 7, Windows Defender Antivirus performed better than the industry average in several areas, and had a significant shortcoming 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. 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 Performance 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 the team's findings, the following improvements were made:

- Improved whitelisting capabilities to save time on scanning known good files
- Built an internal testing system that includes performance gates to identify inefficient operations (from a performance standpoint) and a workflow for remediation

3 Examining the SE Labs results

3.1 Summary of over results

The table below summarizes the overall test results for Windows Defender Antivirus in the October-December 2018 antivirus testing by SE Labs:

Test category	Score	Ranking
Protection accuracy	100%	1st out of 11 (tied with 1 more)
Web downloads	75/75	
Targeted attacks	25/25	
Legitimate software accuracy	100%	1st out of 11 (tied with 4 more)

Table 8. Overall Windows Defender Antivirus test results in the October-December 2018 SE Labs test.

3.2 Understanding Protection Accuracy test scores

SE Labs determines the Protection accuracy scores based on the combined outcome of two tests:

1. Web downloads (75 test cases)
2. Targeted attacks (25 test cases)

SE Labs goes beyond the binary rating (i.e., blocked vs. compromised) in rating protection effectiveness. Instead SE Labs considers the nuances of the interaction between the product and the threat. For example, it issues a different rating for *Blocked (+2 points)* from what is given for *Complete remediation (+1 points)* or *Compromised system (-5 points)*. The other ratings used by SE Labs for both Web downloads and Targeted attacks tests are: *Detected (+1)*, *Neutralized (+1)*, *Persistent neutralization (-2)*. A rating is assigned to each product-threat interaction operation and a combined score is calculated for each product.

Windows Defender Antivirus achieved the following combined score across Web download and the Targeted attacks tests:

Detected	Blocked	Neutralized	Compromised	Protected
100	100	0	0	100

Table 9. Summary of Windows Defender Antivirus scores in the Protection accuracy test

When it comes to the Targeted attacks test, the protection score takes into account the extent of protection demonstrated by the product (i.e., the attack stage in which the product was able to block the threat). Points are deducted for *Access (-1)*, *Action (-1)*, *Escalation (-2)*, and *Post-escalation action (-1)*. Because Windows Defender Antivirus detected or blocked all the targeted attacks in the test, there were no deductions to the score.

3.3 Understanding Legitimate Software Accuracy test scores

SE Labs' Legitimate Software Accuracy test measures the endpoint product's ability to correctly classify legitimate applications. SE Labs assigns ratings based on how the product classifies an object (safe, unknown, not classified, suspicious, unwanted, or malicious) and the level of interaction required of the user (e.g., click, or no interaction required).

SE Labs also takes into consideration the prevalence of the legitimate application to account for the breadth of business impact of incorrectly blocking. This prevalence factor is expressed as a modifier and is multiplied by the interaction rating to determine the product score.

Windows Defender Antivirus correctly classified 100% of legitimate applications as safe.

4 Examining AV-Comparatives results

4.1 Summary of overall AV-Comparatives scores

The table below summarizes overall test results for Windows Defender Antivirus in the March-April 2019 antivirus testing by AV-Comparatives:

	Real-world	Malware
Overall scores for this cycle >>>	99.7/100	99.5/100

Table 10. Windows Defender Antivirus' overall antivirus test results in the March-April 2019 AV-Comparatives Business Security test. AV-Comparatives uses Real-world protection, and Malware protection, test modules.

4.2 Understanding Real-world protection test scores

The below table displays more details on the results of the Real-World Protection test. The results are based on a test set consisting of 389 test cases (such as malicious URLs) tested from the beginning of March till the end of April 2019. The overall business product reports (each covering four months) will be released in July and December 2019.

	March-April
Blocked	99.5% (387/389)
User dependent	2
Compromised	0
Overall Real-world protection rate** (all samples)	99.7% (389/389)
Overall Real-world protection score >>>	**99.7%**
False positives	8

Table 11. Summary of Real world protection scores for the March-April 2019 Business security test

**[Blocked % + (User dependent % / 2)]

The table below shows Windows Defender Antivirus detection rates in Real-World protection tests consistently improving over a one-year period.

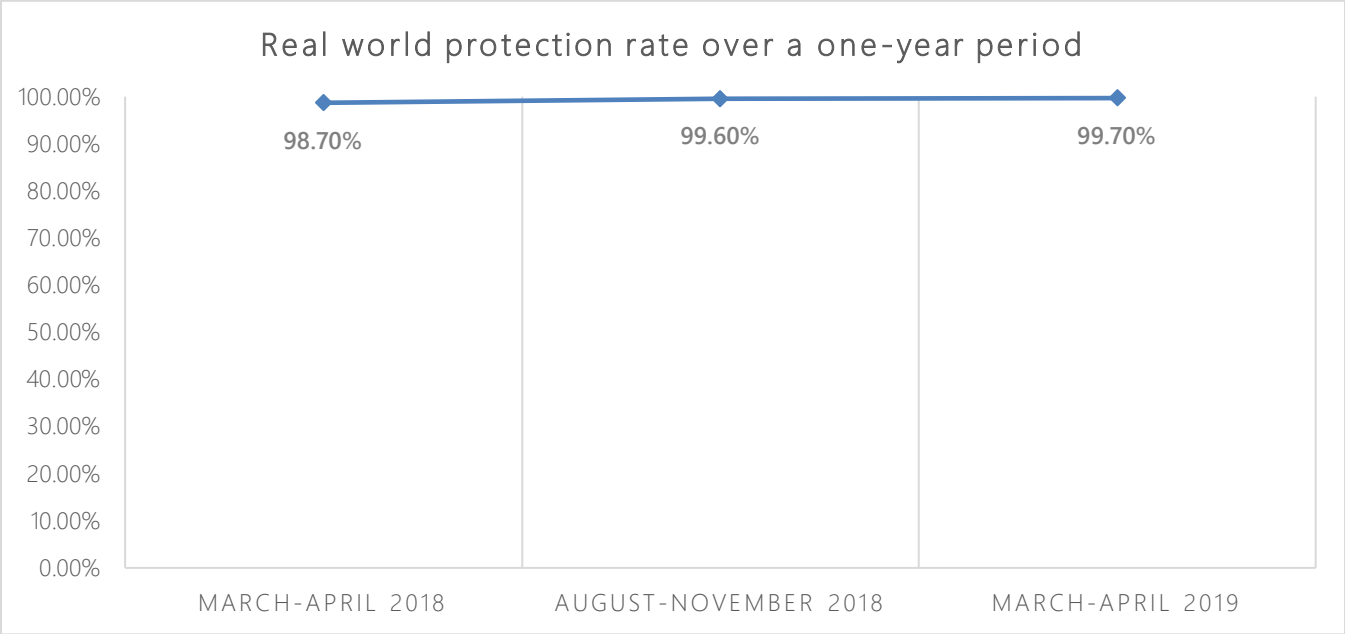

Figure 5. Windows Defender Antivirus detection rates in AV-Comparatives Real-World protection tests over a one-year period

4.3 Understanding Malware protection test scores

The below table gives a brief overview of the results of the Business Malware Protection test run in March 2019. The results are based on a test set consisting of 1311 recent malware samples used during March 2019. Below are details on the Malware Protection test scores.

	March
Blocked	99.5% (1,304/1,311)
User dependent	0
Compromised	0.5%
Overall Malware protection rate (all samples)	99.5% (1,304/1,311)
Overall Malware protection score >>>	**99.5%**
False positives	0

Table 12. Summary of Malware protection scores for the March 2019 Business User test

The table below shows Windows Defender Antivirus detection rates in Malware protection tests over a one-year period. This test is conducted once every six months.

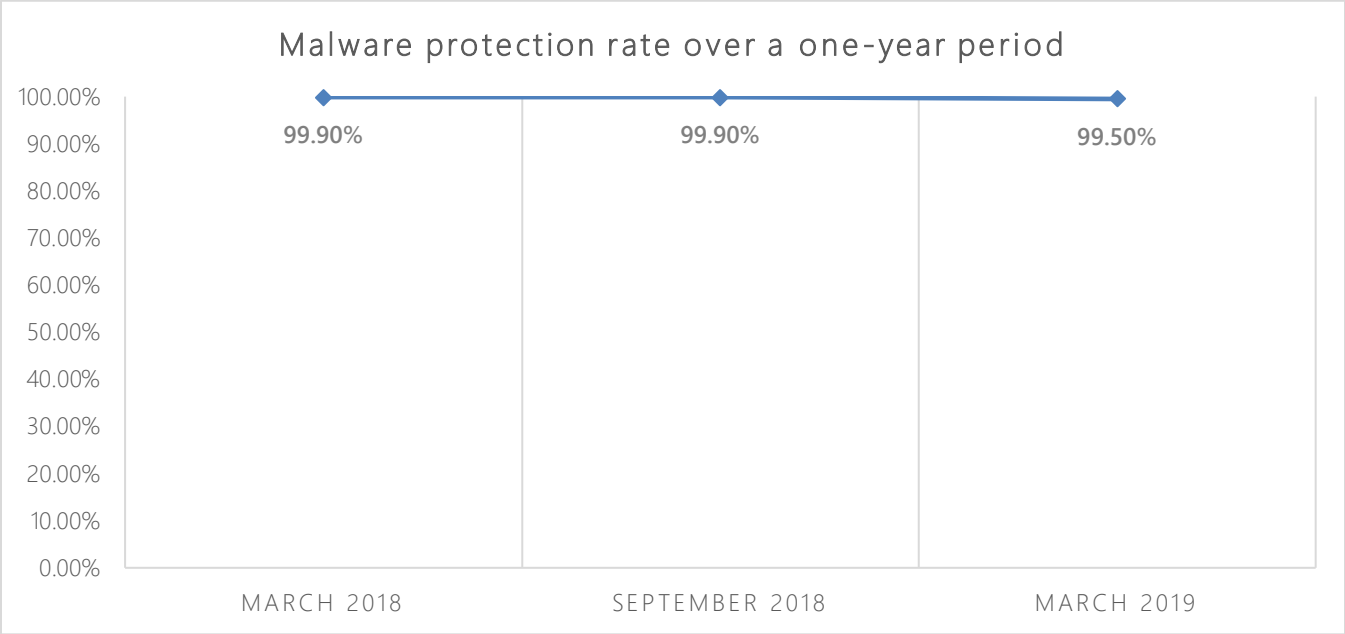

Figure 6. Windows Defender Antivirus Malware Protection rates in AV-Comparatives Malware protection tests over a one-year period

4.4 Analyzing false positives

In the Real-world protection test, Microsoft Defender ATP misclassified 8 files. Below is a list of the files that Windows Defender Antivirus misclassified. As we do for all test results, we analyzed these false positives.

Based on global prevalence data, these files are not common in enterprise environments. All misclassified files are **not** digitally signed. Microsoft encourages software vendors to help minimize false positives by taking steps to raise the level of trust both by security vendors and users.

Sample	Global file prevalence (30 days)	Description	Digitally signed? (Y/N)
Sample a	2	Compiler, customized installer	N
Sample b	2	Tool for renaming JPG files	N
Sample c	50	Hosts file utility	N
Sample d	25	Video encoder library	N
Sample e	846	ISO image downloader	N

Sample f	25	Finance tool	N
Sample g	25	Archiving tool	N
Sample h	76	Backup utility	N

Table 13. Analyzing false positives

As part of the Malware protection test, AV-Comparatives also ran a false positive test with common business software. Windows Defender Antivirus, like all other enterprise security solutions included in the test, had **zero false positives**. This is consistent with our observation about the files that Microsoft Defender Antivirus misclassifies on some tests. Revisit section 2.3.3 for more insights and commentary on false positives.