Microsoft

Transparency report

Examining industry test results, November 2019

Prepared by

Microsoft Defender ATP Research Team

Microsoft

Table of Contents

Microsoft

1 Summary of latest industry test results

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

Over the last few years, Microsoft has been improving its performance in industry tests. Today, it consistently achieves high scores in these tests, demonstrating the strength of our protection capabilities and the innovations we continue to make in our security technologies.

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

1.1 AV-TEST: Perfect Protection scores (July-August 2019)

Windows Defender Antivirus achieved perfect scores (6.0/6.0) in the Protection module of AV-TEST's July-August 2019 Business User test cycle. The industry-leading antivirus solution has consistently achieved this in all AV-TEST cycles in the past 14 months.

In Usability, Windows Defender Antivirus achieved a perfect 6.0/6.0 in July-August.

In the Performance test module, Windows Defender Antivirus achieved a score of 5.5/6.0 in July-August. Learn More >>

1.2 AV-Comparatives: Approved Business Product (August-September 2019)

In July 2019, AV-Comparatives released the Business Security Test August-September 2019 report, which combines results from various reports. Windows Defender Antivirus retained the recognition as an Approved Business Product.

Windows Defender Antivirus achieved a protection rate of 99.9% in the Real-World Protection Test (August-September) and 99.9% in Malware Protection Test (September). Learn More >>

1.3 SE Labs: AAA Award (July–September 2019)

In SE Labs' Enterprise Endpoint Protection test for July - September 2019 (Q3), Windows Defender Antivirus won the AAA Award.

Windows Defender Antivirus registered 98% Protection Accuracy rating and 100% Legitimate Accuracy rating in July – September 2019 test periods for a consistent Total Accuracy rating of 99%. Learn More >>

2 Examining 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 July-August 2019 AV-TEST Business User test:

	July-August
Protection	6.0/6.0 (±0)
Usability	6.0/6.0 (+0.5)
Performance	5.5/6.0 (-0.5)

Table 1. Windows Defender Antivirus' overall antivirus test results in the July-August 2019 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% (368/368)
Prevalent Malware testing	100% (6,572/6,572)
Overall malware protection rate (all samples)	100% (13,521/13,521)
Overall Protection score >>>	**6.0/6.0** (±0)
Overall Protection ranking >>>	1st out of 18 (tied with 14 more)

Table 2. Summary of Protection scores for the July-August 2019 Business User test.

Windows Defender Antivirus detected 100% of malware files used in the Prevalent Malware and Real-World testing in July-August 2019 test cycles from13,889 files used.

The diagrams below show Windows Defender Antivirus detection rates in the Prevalent Malware and Real-World tests 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.

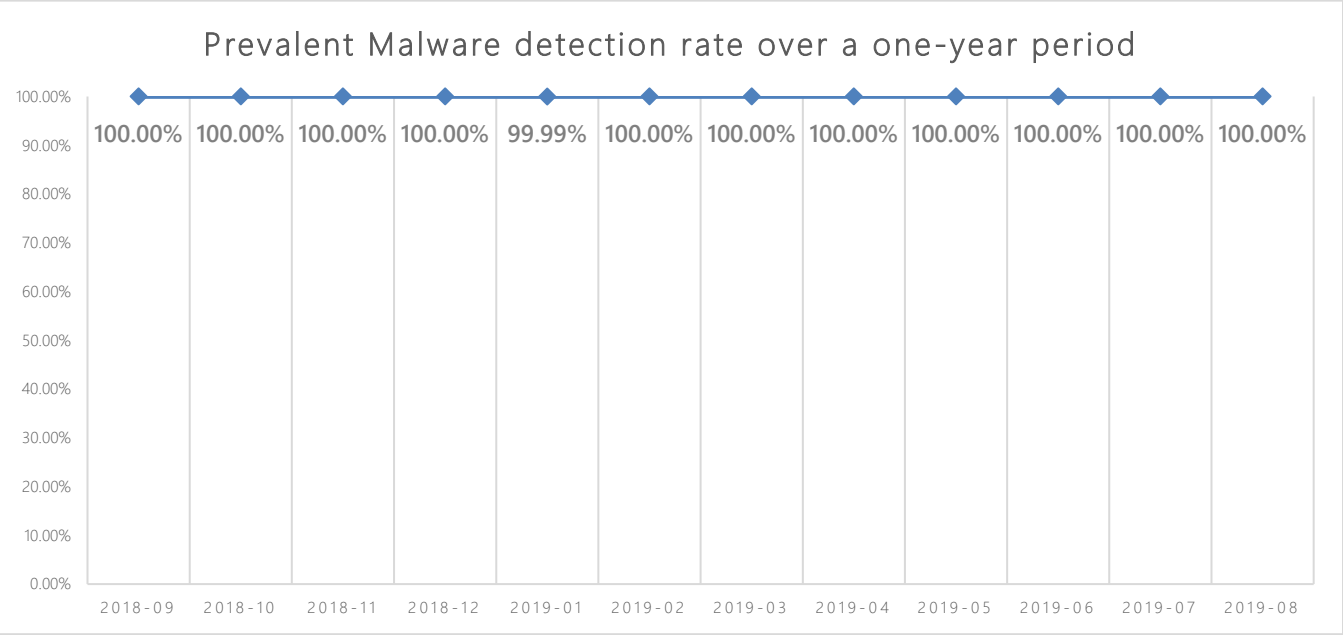

Figure 1. Windows Defender Antivirus detection rates in AV-TEST "Prevalent malware" tests over a one-year period

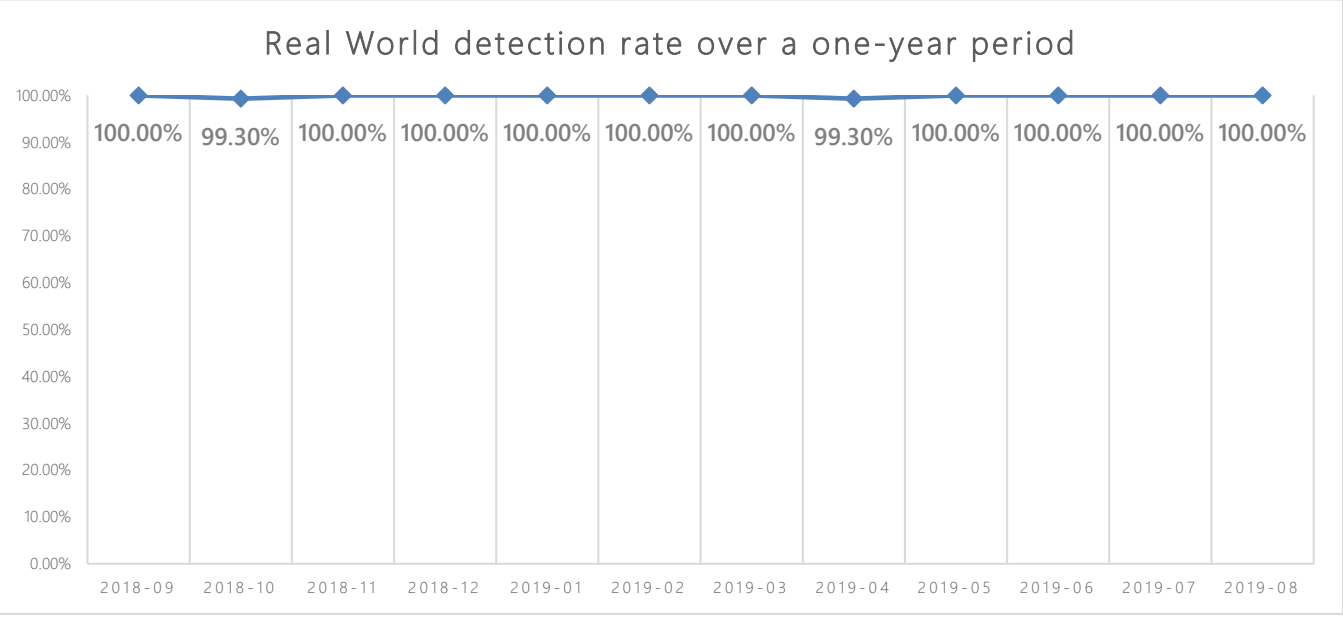

Figure 2. Windows Defender Antivirus detection rates in AV-TEST "Real World" tests over a one-year period

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 the results for Windows Defender Antivirus in the Usability test.

	July	August
Number of misclassified files	0 (out of 14,50,998 samples)	1 (out of 14,50,998 samples)
Overall Usability score >>>	**6.0/6.0** (+0.5)	**6.0/6.0** (+0.5)
Overall Usability ranking >>>	1st out of 18 (tied with 10 more)	1st out of 18 (tied with 10 more)

Table 3. Summary of Usability test scores for the July-August 2019 Business User test

2.3.1 Analysis: What kinds of files were misclassified?

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

Below is an example of a file that Windows Defender Antivirus misclassified in the test cycle. Based on our research and on file prevalence data, the misclassified sample is not common in enterprise environments.

Sample	File prevalence (30 days)	Description	Digitally signed? (Y/N)
Sample 1	100	Standard codec application	N

Table 4. Files that Windows Defender antivirus incorrectly classified as malware during July-August 2019 Business User test

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 is 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.

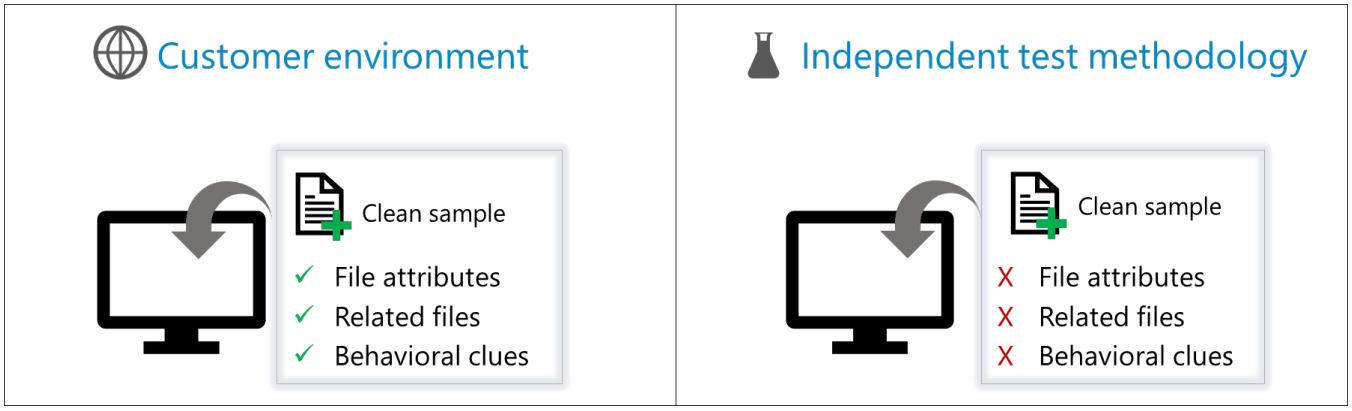

Figure 3. In some cases, samples are incorrectly classified (false positive) in the synthetic test environment but not on customer machines.

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 a 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

2.4 Understanding Performance scores

The table below summarizes Performance test results.

	July-August
Overall Performance test score >>>	**5.5/6.0** (-0.5)
Performance ranking >>>	2nd out of 18 (tied with 6 more)

Table 5. Summary of Performance test scores for the July-August 2019 Business User test

The slight drop in the performance score is due to the higher impact of the product during installation of frequently used applications compared to the last period (from 27% to 36% on standard PCs and 23% to 31% on high-end PCs). Performance continues to be an investment area for the Windows Defender Antivirus team.

The table below presents Windows Defender Antivirus' performance test results compared to industry averages during the July-August test cycle. 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 or the same as the industry average; red boxes indicate performance lower than the industry average.

Operation*	**Standard PC**	*Industry average*	**High-End PC**	*Industry average*
Launching popular websites	9%	*16%*	6%	*15%*
Downloading frequently used applications*	0%	*1%*	0%	*0%*
Launching standard software applications	11%	*12%*	8%	*7%*
Installation of frequently used applications	36%	*24%*	31%	*21%*
Copying of files (locally and in a network)	0%	*3%*	2%	*4%*

Table 6. The average impact of the product on computer speed in daily usage during July-August 2019

*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

Windows Defender Antivirus performed better than the industry average in most areas and had a limitation 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. A 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

3 Examining AV-Comparatives results

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

	Real-World	Malware Protection
Overall scores for this cycle >>>	**99.9%**	**99.9%**

Table 7. Windows Defender Antivirus' overall antivirus test results in the August-September 2019 AV-Comparatives Business Security Test. AV-Comparatives use Real-world protection, and Malware protection, test modules.

3.1 Understanding Real-world protection test scores

The table below presents more details on the results of the Real-World Protection test. The results are based on a test set consisting of 371 test cases (such as malicious URLs) tested from the beginning of August through the end of September 2019.

	August-September
Blocked	370
User dependent	1
Compromised	-
Overall Real-world protection rate** (all samples)	99.9% (369/370)
Overall Real-world protection score >>>>	**99.9%**
False positives	35

Table 8. Summary of Real world protection scores for the August-September 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.

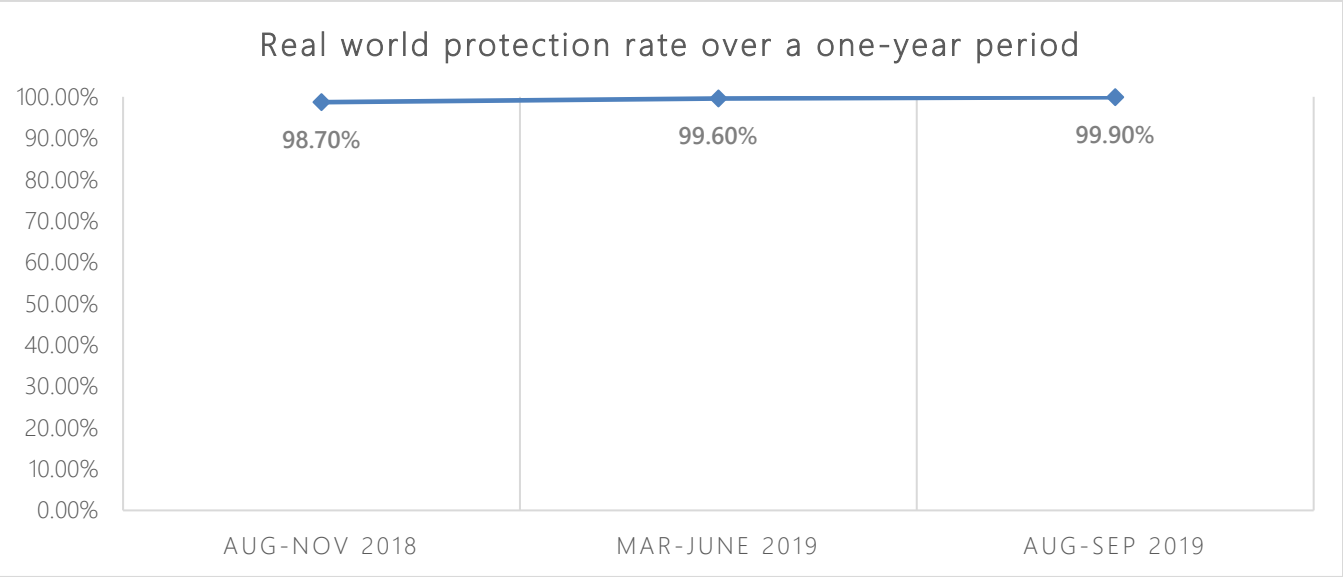

Figure 5. Windows Defender Antivirus detection rates in AV-Comparatives Real-World protection tests

3.2 Understanding Malware protection test scores

The below table gives a brief overview of the results of the Business Malware Protection test run in September 2019. The results are based on a test set consisting of 1,278 recent malware samples used during September 2019. Below are details on the Malware Protection test scores.

	September
Blocked	1,275/1,278
User dependent	0
Compromised	0.5%
Overall Malware protection rate (all samples)	99.7% (1,275/1,278)
Overall Malware protection score >>>	**99.9%**
False positives	35

Table 9. Summary of Malware protection scores for the September 2019 Business Security 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.

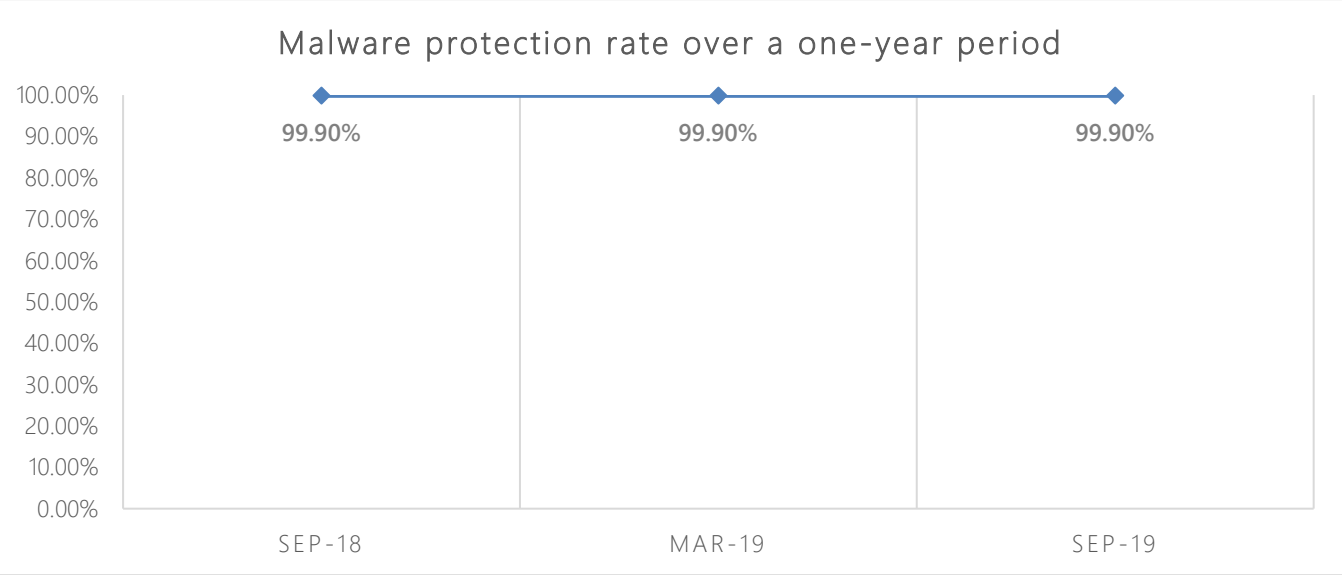

Figure 6. Windows Defender Antivirus Malware Protection rates in AV-Comparatives Malware protection tests

3.3 Analyzing false positives

In the Real-world protection test, Windows Defender Antivirus misclassified 35 files. 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. Most 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. Below is a table with the misclassified files.

Sample	Global file prevalence (30 days)	Description	Digitally signed? (Y/N)
Sample 1	2	Time tracker application	N
Sample 2	2	Disk cleaner tool	N
Sample 3	25	Disk space application	N
Sample 4	2	Photo publish tool	Y
Sample 5	3501	Freeware application for developers for repairing, maintaining, and optimizing Windows computers.	N

Sample	Global file prevalence (30 days)	Description	Digitally signed? (Y/N)
Sample 6	2	Data recovery and email migration tool	N
Sample 7	50	System library files repair tool	N
Sample 8	50	Revision control tool	N
Sample 0	10	Messaging application	N
Sample 10	2	Code grouping tool	N
Sample 11	2	Website conversion application	N
Sample 12	50	Multi-currency accounting software	N
Sample 13	2	Image merger tool	N
Sample 14	2	Web survey tool	N
Sample 15	25	Cloud storage aggregator tool	N
Sample 16	2	Multi-desktop layout saver application	N
Sample 17	25	Duplicate files removal tool	N
Sample 18	3	Files classification & categorization tool	N
Sample 19	25	Video encoder application	N
Sample 20	203	MAC address setup tool	N
Sample 21	2	File synchronization tool	N
Sample 22	1902	Firewall setup application	Y
Sample 23	10	Image slideshow application	N
Sample 24	76	Contact searching tool	N
Sample 25	3	File search software	N

Sample	Global file prevalence (30 days)	Description	Digitally signed? (Y/N)
Sample 26	100	Text editor tool	N
Sample 27	2	Malicious tracking tool	N
Sample 28	11	Unit conversion tool	N
Sample 29	10	Data backup application	N
Sample 30	0	Print application software	N
Sample 31	10	Website conversion tool	N
Sample 32	2	Software deployment tool	N
Sample 33	75	Photo editor application	N
Sample 34	100	Registry finder tool	N
Sample 35	100	Data decompressor tool	N

Table 10. Files that Windows Defender Antivirus incorrectly classified as malware

As part of the Malware protection test, AV-Comparatives also ran a false positive test with common business software. Windows Defender Antivirus 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 for more insights and commentary on false positives.

4 Examining the SE Labs results

4.1 Summary of overall results

The table below summarizes the overall test results for Windows Defender Antivirus in the July-September 2019 testing by SE Labs:

Test category	July-September
Protection Accuracy	98%
Web downloads score	74/75
Targeted attacks score	25/25
Legitimate software accuracy	100%
Total accuracy rating	**99%**

Table 11. Overall Windows Defender Antivirus test results in the SE Labs test.

4.2 Understanding Protection Accuracy test scores

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

1. Web downloads (74 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 a *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 for Web downloads and the Targeted attack tests.

	July-September
Detected	100
Blocked	97
Neutralized	2
Compromised	1
Protected	99

Table 12. Summary of Windows Defender Antivirus scores in the Protection Accuracy test

In the July-September test, Windows Defender Antivirus detected 100 of the samples used and blocked 97. Of the 3 missed samples, 2 were neutralized, while 1 was able to successfully compromise the machine, which resulted in 98% total accuracy rating.

When it comes to the Targeted attacks test, the protection score considers 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)*. **Windows Defender Antivirus detected or blocked all the targeted attacks in the test.**

4.3 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 the 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 in July-September 2019 test cycle.