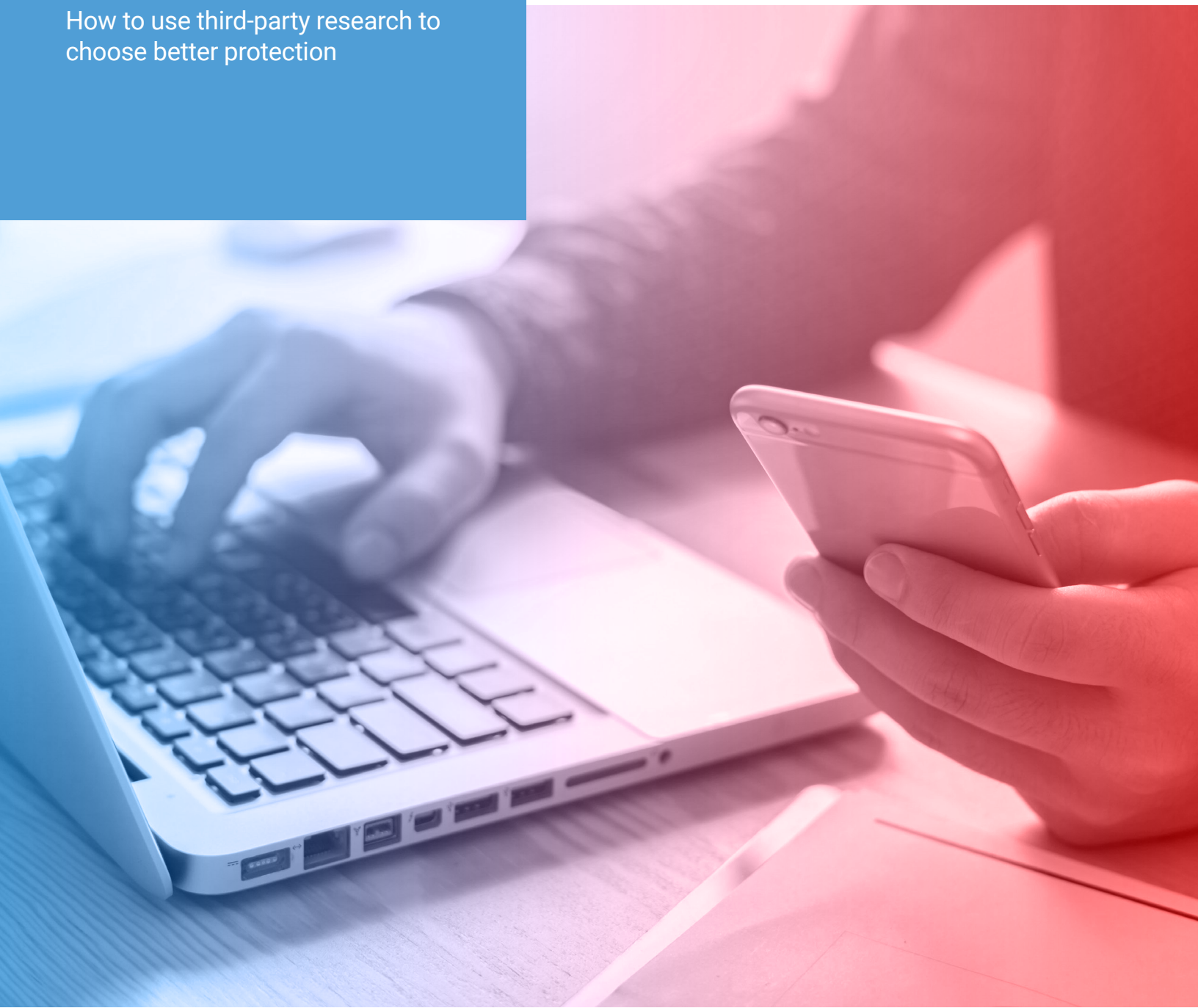
Bitdefender

THE ULTIMATE GUIDE TO AV AND EDR INDEPENDENT TESTS

How to use third-party research to choose better protection



A Broad History of Independent AV Tests

1989
Virus Bulletin begins publishing threat research and analysis

1991
Virus Bulletin hosts first international conference

1999
AV-Comparatives begins as a student project at the University of Innsbruck

2005
AV-Test is founded in Magdeburg, Germany

2007
NSS Labs launches as a startup in Texas, USA

2015
MITRE releases ATT&CK platform

2018
MITRE announces ATT&CK-based evaluations

Whether you're busy running your own business or you're leading a small team of IT professionals, your plate is full. You don't have time to keep up with the nuances, evolutions, and developments happening in cybersecurity every day. You're serving customers, managing device deployments, and looking for the next innovation to push your company ahead in your market! So, how do you ensure that you've implemented the right cybersecurity platform for your business?

You look to the experts - independent, third-party analysts who have real-world experience to understand and evaluate cybersecurity options. In this quick guide, we share a brief history of independent cybersecurity testing, showcase some of the best analysts on the market, and help you edit the common misconceptions about these essential technology educators.

All serious next-gen AV players must compete on the level playing field of independent testing.

Cybercriminals are bypassing AV solutions that rely on signatures to identify threats. Through the explosion of new malware, the rise in fileless attacks, and the commoditization of advanced attack tools, they are changing the rules of the game. And, next-gen AV players exist almost entirely because of the need for cloud-based AI and behavioral technologies that can catch these unknown threats.

Framing the choice as AI-powered versus signature-based AV is, of course, a false premise. Many established vendors have been perfecting AI and using it alongside more traditional technologies quite successfully. The question is instead, how do newer vendors compare to the established vendors? While many newer vendors have been hesitant to engage in public testing, the vital need for establishing objective evidence of their effectiveness has been highlighted by major industry analysts. The 2017 Gartner Magic Quadrant for Endpoint Protection Platforms noted: "Standardized testing, such as AV comparatives and AV tests, are still the best indicators of effectiveness."

Today, endpoint security vendors need to participate in independent tests to be considered serious enterprise-level tools. For example, Gartner includes independent test participation as a requirement for inclusion in its Magic Quadrant for Endpoint Protection Platforms:
"The vendor's nonconsumer EPP must have participated in independent, well-known, public tests for accuracy and effectiveness within the 12 months prior to 30 June 2019, or be a current participant in the VirusTotal public interface. Examples include MITRE ATT&CK Evaluations, Virus Bulletin, AV-TEST, AV-Comparatives, NSS Labs and SE Labs." (1)

(1) Gartner, Magic Quadrant for Endpoint Protection Platforms, Aug 2019

Learn the Language of AV Testing:

Malware merely means "malicious software." From viruses and Trojan horses to ransomware to even adware and spyware, the term includes any intentionally designed coding to cause damage, obtain data without permission, or deceive the end user through misrepresentation.

Public testing is done by third-party labs independently of a vendor's request, sometimes without their involvement at all. These tests typically focus on rigorous standards and issue grades, including failing grades, by category. Some public tests do require a fee to participate, but by charging all vendors equally, they maintain impartiality in the results.

Private or commissioned testing is done at the behest of the vendor and often funded directly by the vendor. While less valuable for customers seeking to compare platforms, these tests provide important validation and feedback to vendors launching new products and services. Furthermore, private tests can provide valuable comparisons against specific vendors that typically avoid public testing, allowing vendors to address marketplace differences in a meaningful context.

Real-world test environments offer the most significant insights into a platform's value. These tests use live malware and viruses and often run for an extended time with continuous live updates of the malicious codes and URLs introduced. They also are designed to select "Allow" anytime user intervention is requested by the platform, ensuring to account for user error in testing the system's true protection capabilities.

Protection is measured by how well the platform fends off malware attacks, which can be reported as the number of "false negatives" or pings that get past the AV's safeguards. The standard varies among tests but is usually a low percentage allowed as the earliest indicator of a platform's integrity.

False Positives are clean files that are improperly identified as malicious. This standard is also measured as a low percentage to receive a passing grade and is an essential indicator of the platform's accuracy.

Performance takes stock of how resource-intensive each platform is on the system environment and is an essential consideration in balancing the degree of a platform's protection against its drain on the system's capabilities.

So, what exactly IS an Independent Cybersecurity Test?

With the average cost of a data breach reaching $3.86 million dollars, and the cost of a "mega breach" (>1 million records compromised) soaring by a factor of 100x, cybersecurity is not something anyone can afford to gamble on. And, with an average time to identify and contain a breach stretching out to 280 days, the countdown to your own risk of exposure has possibly already started.[(2)]

While you are likely already quite aware of what's at stake, it is not often easy to understand how the dozens of platforms available on the market compare, or even how they stack up against their own claims. You need to KNOW that your investment is going to work. And, it's unlikely you have the time, expertise, and risk tolerance to introduce a live threat into your own system for thorough and objective testing.

Thankfully, several impartial, third-party organizations have stepped into this knowledge gap to help. Some have been around for more than a decade, while others are relatively recent additions to the industry. They have all evolved their test methodologies alongside the industry. While they each use a different scoring system (and not every test includes every platform), they generally share some level of rigorous standards, a degree of transparency in their methodology, and a commitment to impartiality among vendors.

Who's testing the best of the best?

Publicly conducted tests performed by independent third-party labs are the gold standard in today's crowded cybersecurity marketplace. These tests share several common traits:

- Vendors do not fund them
- They test performance, false positives, and protection
- They pull from massive collections of malware
- They offer vendors the chance to dispute or clarify results
- They do fail platforms in one or more categories based on results

Based on these criteria, here are some of today's leading independent, third-party, antivirus and antimalware labs.

AV-Comparatives

At a Glance

- Certified by the European Institute for Computer Anti-Virus Research (EICAR)
- IS0 9001:2015-certified by TÜV Austria
- Audited by TÜV Austria every year for quality-control

AV-Comparatives is known for their Real-World Protection, Malware, and Performance tests. Given that many of today's threat vectors originate online, their Real-World Protection test is an especially critical analysis of any antivirus or antimalware platform for Enterprise customers protecting vital business data and assets.

(2) IBM, Cost of a Data Breach Report, 2020

How do AV tests work?

While each test has its nuances, generally they each follow a similar process for a cohort of selected platforms:

1) **Install the platform on the test environment with default settings - most tests run on a Windows platform, but specific tests for Mac, Android, and Linux are also conducted in some tests.**
2) **Introduce the threat vectors - a sampling of thousands of files with known percentages of corrupt and malicious files from carefully maintained databases.**
3) **Conduct the test over a period of time - evaluating for false positives and false negatives.**
4) **Report results to each vendor, allowing time for questions and disputes.**
5) **Publish results to the public.**

The AMTSO Testing Protocol Standard

In 2018, the Anti-Malware Testing Standards Organization (AMTSO) created "an agreed set of rules for transparency, balance, and clarity" among testing labs.

1) **Testing must not endanger the public [by creating new malware].**
2) **Testing must be unbiased.**
3) **Testing should be reasonably open and transparent.**
4) **The effectiveness and performance of anti-malware products must be measured in a balanced way.**
5) **Testers must take reasonable care to validate whether test samples or test cases are accurately classified as malicious, innocent, or invalid.**
6) **The testing methodology must be consistent with the testing purpose.**
7) **The conclusions of a test must be based on the test results.**
8) **Test results should be statistically valid.**
9) **Vendors, testers, and publishers must have an active contact point for testing-related correspondence.**

(2) https://www.scmagazine.com/home/security-news/organization-creates-anti-virus-testing-best-practices/

Regarding their Real-World Protection tests, AV-Comparatives says, "These tests evaluate the suite's 'real-world' protection capabilities with default settings (including on-execution protection features). We aim to do these tests rigorously. Due to that, these tests are time and resource expensive, so only products chosen for the yearly main test-series are included."

Essentially, AV-Comparatives' Real-World Protection test focuses nearly exclusively on protection rather than false positives or performance. This test also does not consider USB flash drives or LAN threats, only those originating via the internet.

The false-alarm test in the Whole-Product Dynamic "Real-World" Protection Test consists of two parts: wrongly blocked domains (while browsing) and mistakenly blocked files (while downloading/installing). It is necessary to test both scenarios because trying only one of the two above cases could penalize products that focus mainly on one type of protection method, either URL filtering or on-access/behavior/reputation-based file protection.

AV-Comparatives also offers an Enhanced Real-World Protection test, introduced in December 2019. As attacks continue to grow more sophisticated, testing must also grow more thorough. The first round of this new test selected a subset of vendors to run against advanced attack scenarios, testing not only their ability to stop the attacks, but when they stopped each attack, thus evaluating their ability to stop data breaches before attacks execute. (3)

AV-TEST

At a Glance

- An independent lab headquartered in Germany
- Founded by Andreas Marx
- Releases monthly test results to the public

AV-Test focuses on certifications and is known for their fluid ranking system that evaluates antivirus software, antimalware tools, and security software for Windows, Mac, and Android platforms. Of particular note is their malware test, running continuously against a database of more than 3 million potentially malicious files, websites, and emails, with up to 35,000 new threats added daily. Each year, AV-Test presents their Best Protection Award to one platform in each of their testing categories for both business and consumer audiences. AV-Test evaluates each cybersecurity platform using in-house analysis software against three criteria: protection, performance, and usability. Vendors can earn 6 points per category, placing a perfect score at 18 points. Any vendor that successfully passes a full testing cycle receives the AV-Test Certification & Approval for one year. To see how a given vendor has performed over history, it's as simple clicking on the vendor's name and viewing their full history of scores for any given category of testing. (4)

(3) https://www.av-comparatives.org/test-methods/
(4) https://www.av-test.org/en

MITRE

At a Glance

- Developed the ATT&CK knowledge base, providing free access to "real-world reporting of adversary tactics and techniques."
- Operates in the United States as a public-private partnership through FFRDCs (federally funded research and development centers).

As a publicly available resource, the ATT&CK platform launched in 2015 and quickly gained a reputation as a valuable asset for cybersecurity professionals, focusing on testing and evaluation by internal teams coordinating between red teams, defenders, and managers.

The MITRE ATT&CK Matrix for Enterprises provides an approachable map of tactics and techniques observed during cyberattacks, a resource many security vendors have since adopted. In fact, many vendors now consider the ATT&CK platform's terminology as an industry standard in describing cyberattacks.

In 2018, MITRE announced its first cohort of ATT&CK-based product evaluations, which have rapidly become the standard in comparing Endpoint Detection and Response products. All results are released to the public, in line with MITRE's mission of "providing objective insight." However, it's worth noting that the evaluations are not certifications but rather independent analyses of each platform's detection capabilities. The results are not scored, ranked, or rated, nor are platforms compared against each other.

By using what MITRE calls "adversary emulation" or testing "in the style of a specific adversary," the ATT&CK evaluations focus the test on subsets of established techniques. Most recently, MITRE APT 29 tested products against the emulated APT 29 threat group. For each vendor tested, threats detected are noted along each step of the attack. This evaluation distinguishes between detection types: Telemetry and MSSP detections are inherently delayed and often difficult for the average enterprise IT department to act on without the help of an SOC team. Whereas, General, Tactic, and Technique categories identify suspicious action, determine what the attacker is trying to achieve (such as gaining persistence), as well as how they are attempting to achieve it. [5a] [5b]

NSS LABS

At a Glance

- Launched as a startup company in Texas to supplement available third-party testing
- Known for advanced endpoint protection tests with the inclusion of Total Cost of Ownership calculations for every solution tested

As one of the newer testing labs, NSS focuses on threat and vulnerability research, seeking to understand "the many different ways attackers can circumvent security products." As a result, they are well equipped to measure a cybersecurity platform's effectiveness and performance, even examining stability, usability, and ownership cost.

Their testing methodologies are based on complimentary public briefings with users, analysts, and vendors. Based on insights gathered in these recurring dialogues, the company updates its procedures, which are then vetted by an advisory board. As products are evaluated, they're rated as either Recommended, Neutral, or Caution. It's important to note that all products start with a Caution rating and earn their final rating based solely on test results and empirical data.

Vendors are invited to participate in our group tests based on their market presence or at the request of enterprise customers. While participation in an NSS group test is always free, they defray their costs by selling the test results to interested customers through a subscription and selling marketing rights to interested vendors after testing is completed. [6]

(5a) https://www.mitre.org/news/press-releases/mitre-releases-results-of-evaluations-of-21-cybers ecurity-products
(5b) https://attackevals.mitre-engenuity
(6) https://www.nsslabs.com/about/

Virus Bulletin

At A Glance

- Known for their annual conference and bimonthly certifications
- UK-based testing lab with decades of experience

Having been in business for over 30 years, Virus Bulletin is known for its VB100 award. For many years, this de facto certification has been considered a minimum standard of quality for malware detection due to its simplicity: "proving whether a product can detect 100% of malware samples listed on the WildList without generating any false positives."

Virus Bulletin is not a comparative performance analysis across platforms, nor is it real-world in its methodologies. However, it is a static, first-line standard for establishing detection capabilities equally across any cybersecurity platform in the world.

The VB100 Certification Test uses a "quasi-standard body of samples, changing from test to test" to measure performance via exposure to a known set of malware samples for several days in both on-access and on-demand modes. Typically, these tests include the WildList and the AMTSO RTTL lists.

The ensuing Diversity Test exposes the platform to a broader selection of samples. However, this is not considered part of the certification process and is merely complementary data for contextual purposes. The VB100 test also determines false positives using a clean set of more than 400,000 files maintained by VB. (7)

How do I know which one to trust?

First of all, let's get rid of a common misconception - that antivirus tests are "pay to play" advertisements in disguise. While there are undoubtedly numerous "certifications" that can be bought, the independent tests covered in this whitepaper stand up well to scrutiny. If this were reality, the results would tell the tale pretty quickly because the platforms with the most significant budgets would always win the most recognition! The fact that you can look across time and the spectrum of available types of testing and consistently find small cybersecurity innovators acknowledged for their contributions to the field should prove that these tests are more about validation than monetization.

Here are three key tips that will help you understand how to evaluate any antivirus test's validity, even if it's not covered in this whitepaper.

Follow the Money - The funding for the top independent AV and EDR tests is transparent and is worth considering when assessing the reliability of the results. As a general rule, private tests, sponsored by a single vendor, tend to be considered less reliable than the ones where vendors equally contribute or funding is provided from other sources.

Look for Trends - With enough tests, particularly private commissioned tests, any product can achieve a great result once. Instead, look beyond a single snapshot or moment in time and evaluate the options against a wider range of tests that establish a trend of proven, consistent performance over a longer duration of time.

Don't Trust Just One - if you look to a single test to confirm your purchase, you are likely missing nuances. Instead, look for confirmation across multiple independent tests before making your decision. As we've shown above, each test emphasizes different strengths, weaknesses, and manners of reporting. This takes us back to where we started - you know your business better than anyone else. Now that you have this array of deeply researched evaluations at your fingertips, you are much better equipped to make decisions about the right cybersecurity partner to protect your digital assets. Although there is no perfect test, nor will there likely ever be, independent testing for software platforms is a valuable indicator of real-world protection effectiveness and a powerful tool for busy IT leaders who need trustworthy validation of the protection they choose for their company.

(7) https://www.virusbulletin.com/testing/vb100/vb100-methodology/vb100-methodology-ver1-1/

How has Bitdefender performed on independent AV tests?

Bitdefender endpoint security consistently ranks first for protection or detection in AV and EDR tests from the most reputable independent testing organizations. From June 2016 to present day Bitdefender has achieved a maximum score of 6 for protection in all 26 AV-TEST business trials. Bitdefender has also dominated the 2019 AV-Comparatives enhanced detection test with perfect protection at pre-execution and has been the only vendor to achieve a perfect score against advanced threats in 2020. In the MITRE APT 29 evaluation published in April 2020, Bitdefender demonstrated the highest number of contextual detections and achieved full attack coverage with actionable detections.

Bitdefender Security for Mail Servers solution, powered by the antispam technology, is the only product to have received a certification in all VBspam tests ever performed and the only solution to have won 24 consecutives VBSpam+ awards, the highest certification awarded in the VBSpam Tests performed by Virus Bulletin.

Learn more about independent testing and the AI and behavioral technologies that give Bitdefender the edge.