

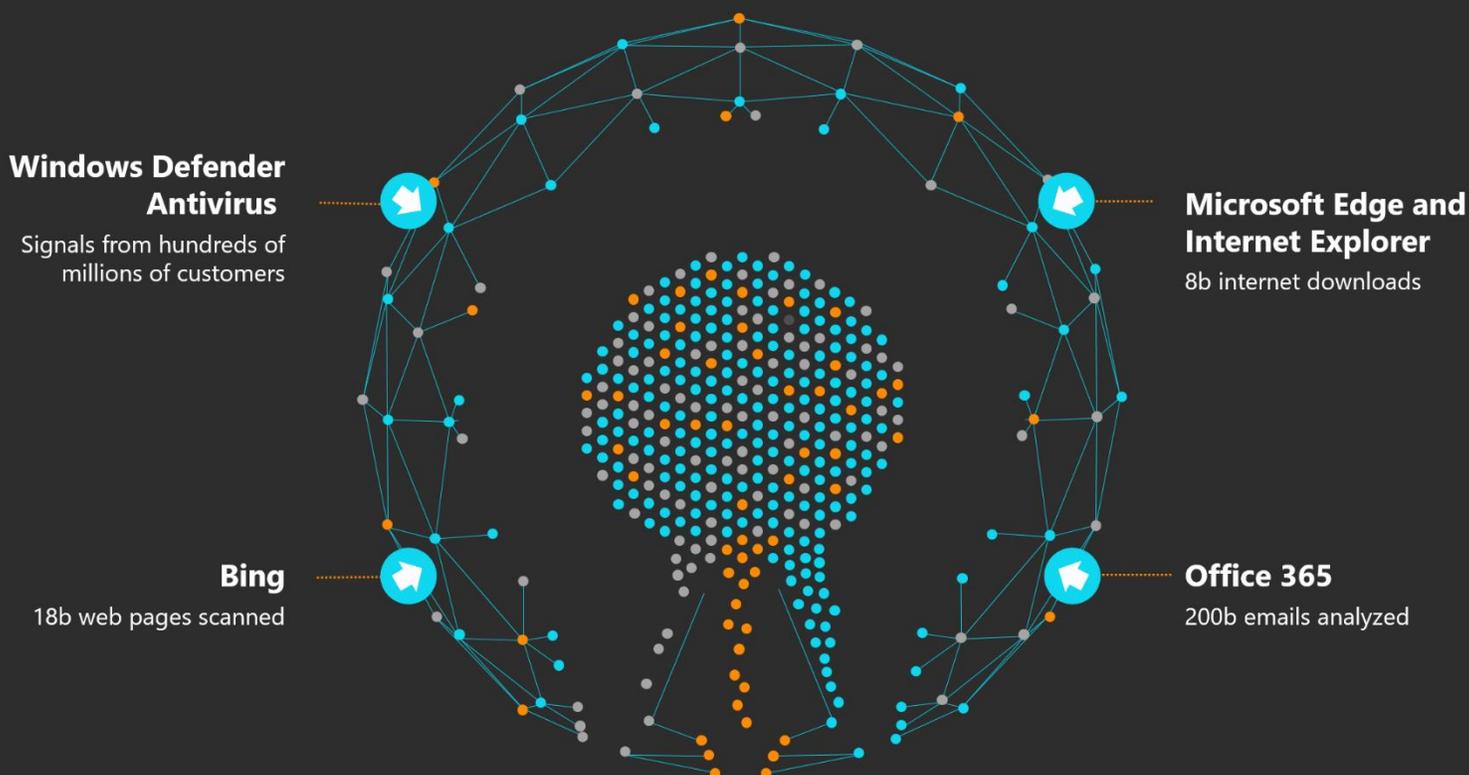
# Evolution of malware prevention

In a mobile-first, cloud-first world, people stay productive and connected using a variety of devices. While there is incredible value in so much connectivity and productivity, there is a corresponding growth in risk as people increase their exposure to cybersecurity threats. While security has always been a priority for Microsoft, this new world requires a new approach to and a [large investment](#) in threat prevention, detection, and response. [Windows Defender Antivirus](#), along with many other [features](#) that are built into Windows 10, are at the frontlines and must constantly evolve to protect customers against today's threats and those that will emerge tomorrow.

Traditional, signature-based approaches to malware detection simply do not scale to the cyberthreats customers face. Next-generation detection solutions, like Windows Defender Antivirus, protect customers through data science, machine learning, automation, and behavioral analysis that are guided by expert threat researchers. These next-gen methodologies are required to deliver effective threat prevention in an era of unprecedented attacker activity.

Data science and machine learning has long been a pivotal component of Microsoft. Microsoft Azure and PowerBI are examples of the kinds of products that enable customers to empower themselves with data to gain actionable insights. Behind the scenes of these products is a powerful cloud infrastructure for big data and machine learning algorithms.

Here at Microsoft, we work across product groups to help shape and co-develop these kinds of technologies by using them on real-world problems and experiments. This collaboration helps us build highly efficient machine learning algorithms that are integrated into our client and cloud protection systems, helping us scale and protect a over a billion devices using Windows Defender Antivirus technologies.



Microsoft also has a unique ability to correlate signals from vast domains, such as consumer and corporate email services, online search, and web browsing, on top of malicious and suspicious signals. These signals are collectively processed to deliver protection through Windows Defender Antivirus and Windows Defender Advanced Threat Protection (ATP) either locally or through their cloud services. Combined, these domains leverage threat data from over a billion devices, 18 billion search result pages scanned by Bing, 300 billion authentications, and 200 billion emails scanned for malicious content each month.

Microsoft's unique insights into the threat landscape, informed by trillions of signals from billions of sources, create an [Intelligent Security Graph \(ISG\)](#) that we use to shape how we protect all endpoints, better detect never-before-seen attacks, and accelerate our response. The Intelligent Security Graph is powered by inputs we receive across our endpoints, consumer services, commercial services and on-premises technologies. All that uniquely positions us to personalize our protection and identify anomalies that often represent new threats.

## Our approach

In the modern and connected world, people encounter cyber threats every day, and Windows Defender Antivirus customers are no different. Although some of the malware our customers encounter have been seen before, most of the time it is unique. Blocking at first sight, without having prior knowledge of a specific malicious pattern, is critical. Only next-generation solutions that use data science, machine learning, automation, and behavioral analysis are capable of blocking malware at first sight.

Although some prevalent malware can attempt to infect tens of thousands of customers, it's more likely that a new malicious file will affect very few. In fact, 96% of all malware seen in the first quarter of 2017 was only seen once and blocked at first sight on that single computer.

What is behind this predictive ability to block at first sight? Here are some of our techniques:

- Lightweight, client-based machine learning models block attacks and flag suspicious activity for additional analysis by the cloud protection system.
- Computationally-intensive cloud-based machine learning models deliver verdicts based on signals sent from the client within milliseconds, but can also request a file for additional analysis and return a block or allow verdict to the client within seconds.
- Local behavioral analysis tracks malicious actions in memory and across processes to stop file-based and file-less attacks.
- High-precision "traditional" antivirus on the client efficiently detects common malware, often through generic or heuristic methodologies, and exclude common clean programs from unnecessary scanning and performance impact.

These techniques are implemented on the client and in our cloud protection system. Our protection client often blocks or allows activity based on local classifiers, heuristics, and behavioral or contextual clues. This analysis happens instantly and allows the client to block 97% of the malicious activity our customers encounter. If the malicious intent remains questionable after assessment at the endpoint, a query containing rich metadata is sent to the Windows Defender Antivirus cloud protection system. Numerous models assess the current attack activity while combining data from our global network of protected clients and our Intelligent Security Graph, which correlates threat activity across all our services.

The cloud protection system usually issues a verdict within milliseconds based only on this metadata. If this lightweight cloud analysis is insufficient to reach a determination, the sample is requested for deeper analysis and further processing. This fully-automated deep analysis delivers precise assessments back to the client, but also has the benefit of providing immediate protection against similar threats on devices around the world based on the automated analysis of that one sample. This analysis process and the protection it generates takes only seconds to complete.

A recent example from a real customer illustrates the end-to-end process.

A customer in Texas running Windows 10 Home was chatting on the Discord app and was tricked into downloading a zero-day keylogger belonging to an entirely new malware family. The customer clicked on a link that pointed to a file named *csrss.exe*. As soon as the web browser downloaded the file, Windows Defender Antivirus started scanning it.

In this case, the client-based analysis didn't return a confirmed malware verdict, but the analysis signals were suspicious enough that Windows Defender Antivirus temporarily prevented the file from running until it could request analysis from the cloud protection system. Windows Defender Antivirus sent a query to the cloud protection system, which ran 167 cloud models against the query and returned an initial answer in 435 milliseconds. Still, the answer wasn't definitive—the particular file and contextual metadata around the incident was suspicious, but the final answer from machine learning models leveraging the signal metadata wasn't conclusive enough to block the file. The cloud requested a hold on the sample so that it could upload the file and do further automated analysis. In this case, the client's default settings were set to allow a 10-second delay for this extra processing, which needs to happen only when Windows Defender Antivirus cloud protection system encounters a suspicious signal for the very first time.

As soon as the sample was uploaded, extensive data was extracted from the file and a multiclass Deep Neural Network (DNN) Machine Learning classifier ran on the extracted data and rendered a verdict of malware. Seven seconds after the hold and sample request, the cloud protection system sent a block request to the client, which then replied with a successful block. After this initial assessment, all other customers using Windows Defender Antivirus that encountered that same file and other files with similar properties, were instantly protected from that point forward.

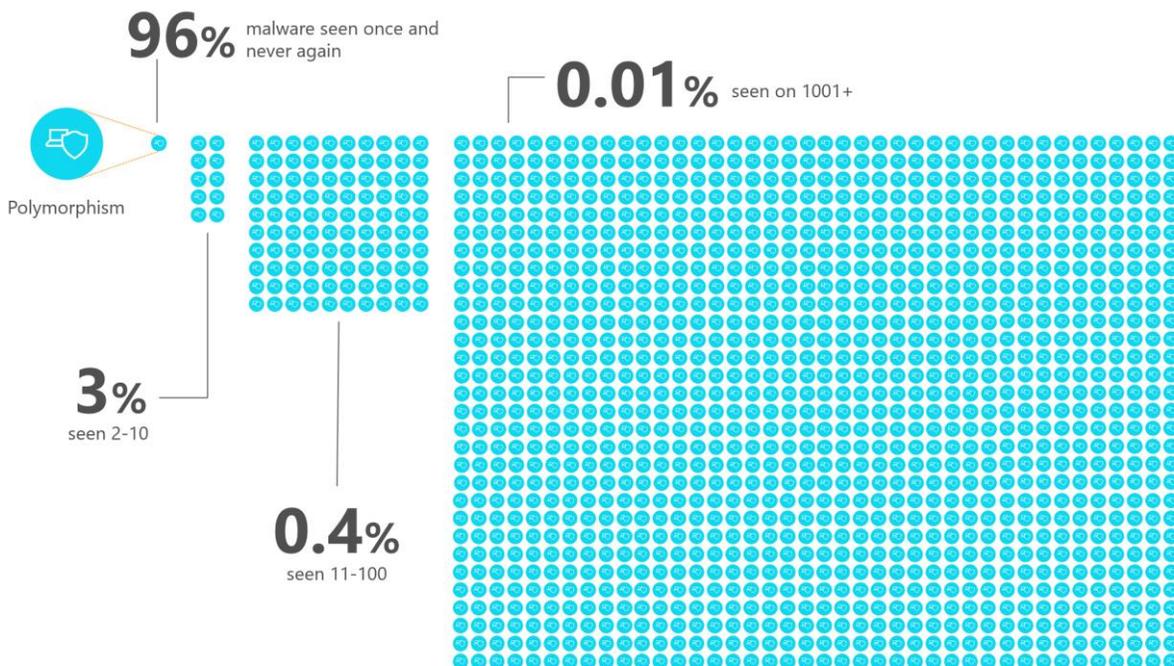


Due to these capabilities, only a few of the most unique files and signals need to be analyzed by our researchers. These files, URLs, and other signals or samples are sometimes the result of incorrect evaluations, such as missed detections or false positives, but they are often borderline cases. Borderline cases are software, domains, or other signal components that walk the line between good and bad, requiring direct verification against our detection criteria. These issues reach us as customer submissions, but they are also flagged by automated systems that surface the most unique and impactful samples and signals, which can better tune our automation and machine learning systems. Researchers and data scientists leverage these expertly-labeled samples along with the existing labeled set to define fresh rules, enhance automation, and train new and existing models, continuously evolving protection.

## Why take a new approach

A look at the data on the threat landscape tells us why a new approach is needed. Let's first talk about scale. Windows Defender Antivirus customers experience around 90 billion potentially malicious encounters per day that need a verdict. *Does an activity represent something malicious? Or is it benign?* On any given day, around 97% of these verdicts are made by the client. The remaining 3% of these encounters, around 2-3 billion queries per day, are processed by the Windows Defender Antivirus cloud protection system. While many decisions can be made on metadata in the query alone, a small percentage of samples are requested for further processing and automation. Along with data from industry partners, we process around 4.5 million files and data points per day through our automated systems. Traditional processing and signature generation simply could not scale to cover sheer number of encounters.

On top of sheer scale, we face the issue of polymorphism. Attackers use modern infrastructure and cloud capabilities to continually generate new threats and package threats in new ways. Most customers encounter attacks that are completely unique to them. Only the most prevalent 1% of malware are ever seen on more than 10 different computers. In fact, the majority of malware—96%—are seen on only one computer and never seen again.

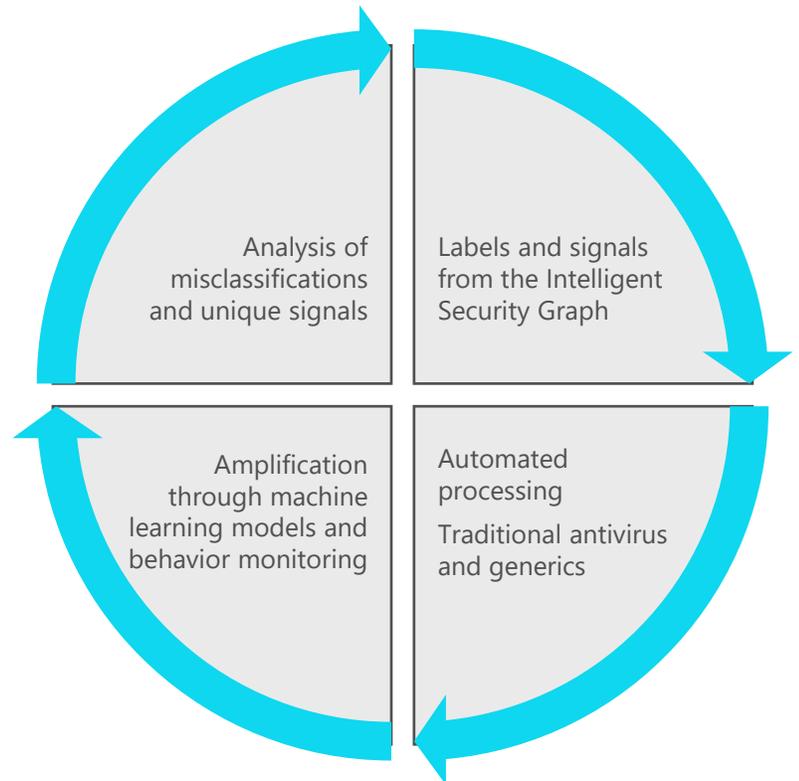


Traditional, signature-based antivirus doesn't predict. It can only make exact or, at best, fuzzy matches to threats that have already been seen before. It is reactive by nature. It's imperative that next-generation antivirus systems can instantly analyze and predict an attack at first sight—possibly the only time that a threat will ever be seen. Expert systems, like machine learning models, must exponentially amplify protection from a limited number of samples to protect customers from millions of never-before-seen malware.

Currently, for every sample analyzed by a Microsoft expert, we protect against an average of 4,500 other malicious samples through our next generation antivirus technologies.

Traditional methods of analyzing and detecting can't scale, but that doesn't mean human analysis isn't important. On the contrary, next generation antivirus protection relies on accurate labels from expert analysis to accurately train performant models. Microsoft's approach leverages this expert analysis along with all our data from the Microsoft Security Intelligence Graph to amplify protection through machine learning, automation, and behavioral analysis, which are delivered through client and cloud-based protection.

On average, a manual investigation amplifies protection against other threats to 12,000 customers.



# Windows 10 prevention, detection, and response

Scale and polymorphism demand a shift in antimalware protection, and Windows Defender Antivirus has been designed specifically to address this extremely difficult and high velocity threat landscape.

Billions of threat signals in the Intelligent Security Graph, our cloud computing power, and our experts in data science, machine learning, and threat research provide us with an unrivaled potential to protect our customers. Much of that potential has already been fully realized with Windows Defender Antivirus on Windows 10.

Of course, Windows Defender Antivirus is just one key component in the fight against malware and other types of threats. Windows 10 includes a stack of security features that complement Windows Defender Antivirus. We've recently introduced [Windows Defender Advanced Threat Protection](#) (Windows Defender ATP) to the Windows Defender brand family, which can help customers to detect and respond to advanced attacks that might get past your primary defenses. These features combined provide a secure and full-featured suite of solutions to help customers achieve the security profile that today's modern threat landscape and customer demand.

<b>Traditional and modern malware defense built-in to Windows Defender Antivirus</b>	<b>Windows Defender Machine Learning-based cloud protection</b>	<b>Windows 10 application isolation, control, exploit mitigation</b>	<b>Rich investigation experience with Windows Defender ATP</b>
<p>High-precision, traditional antivirus for efficient detection of pervasive threats</p> <p>Evolved threat intelligence powered by predictive machine learning, behavioral analysis, and heuristics for new and evasive threats</p> <p>Traditional, behavioral, and efficient next-gen models detect 97% of threats on the client.</p> 	<p>Intensive, machine learning models, dynamic file analysis and other intel-based features are powered by Windows Defender cloud protection system</p> <p>Fused with the Microsoft Intelligent Security graph, directs clients to act within seconds.</p> 	<p>Windows Device Guard locks down system to run only trusted applications</p> <p>Windows Application Guard isolates threats to an application's container</p> <p>Enhanced Mitigation Experience Toolkit (EMET), Hyper-visor Code Integrity (HVCI), and Control Flow Guard (CFG)</p> 	<p>Breach detection, remediation and response</p> <p>Understand your current threat landscape, explore 6 months of rich machine timeline that unifies security events from Windows Defender ATP, Windows Defender Antivirus, and Device Guard and take immediate actions</p> 