

# HOW TO HARNESS MACHINE LEARNING

Tap into the Webroot DNA  
for Security at Scale

WEBROOT®



# MACHINE LEARNING DNA

Powerful threat intelligence powered by machine learning can help process the enormous volume of data that is overwhelming traditional IT security infrastructure and human researchers. The right threat intelligence integrated into devices can reliably block a majority of threats automatically, freeing up humans to focus on the few (most dangerous) threats that get through. Artificial intelligence (AI) is about emulating human intelligence. Machine learning sorts through and analyzes large quantities of data at the speed and volume that is simply impossible for humans alone to process.

*But not all machine learning technologies are created equal.*

We'll investigate five of the most pressing issues facing cybersecurity professionals today and how Webroot's 6th generation machine learning can help stop these sophisticated attacks today and in the future. After all, machine learning is in the Webroot DNA and we've been working with the technology for over a decade.

# TODAY'S SECURITY SOLUTIONS ARE NOT KEEPING PACE



**200%**

increase in attacks from 2016 to 2017. 2018 has already surpassed 2017's numbers.

*Source: CVE Details*



**60%**

increase in Phishing in the first half of 2018

*Source: Webroot, F5 Labs 2018 Phishing and Fraud Report*



**68%**

of phishing sites now use HTTPS to better impersonate legitimate sites which always use HTTPS.

*Source: Webroot, F5 Labs 2018 Phishing and Fraud Report)and Fraud Report*



**2X**

IoT connected devices install base worldwide in the last five years. Projected to grow to 31 billion.

*Source: Statistica*



**94%**

of malicious executables are polymorphic, created to circumvent traditional antimalware solutions.

*Source: 2018 Webroot Threat Report*

# TODAY'S THREAT LANDSCAPE

## Information security today is a data flood that's getting deeper.

Organizations are inundated with an avalanche of information flowing in, out, and across the corporate network in real time—each data point a potential clue into an impending attack. Miss the needle in a haystack, and malicious actors suddenly have access to sensitive data. Just one slip up can result in significant financial loss, a loss of public confidence, and significant PR headaches that can put a company out of business.

## Attacks aren't slowing. And today's threat landscape is building at a sophisticated, persistent pace.

People want to be praised. They want to be entertained. Everyone has a fear of missing out. And because of this, people are trusting. Too trusting. Unfortunately, traditional security infrastructure hasn't evolved enough to protect users from themselves. As a result, today's front lines are being fought by users—untrained in cybersecurity tactics and intelligence, just trying to do their jobs. In fact, 90 percent of successful network breaches are the result of user error.

*It can happen to anyone...*



## Major Headaches for Health Clinic

### Victim Profile

- 40+ year-old neurology clinic
- \$75 million annual revenue
- 40-45 employees

While investigating a ransomware attack, the clinic's IT team discovered a separate data breach that exposed patient records—including names, Social Security numbers, driver's license, addresses, phone numbers, medical data, prescriptions, and insurance data—to an attacker for 15 months.

Up to **400K** patient records may have been compromised. The average healthcare data breach cost \$380 per record in 2017. Let's do the math.

**Loss: Up to \$152 Million**

# EXISTING SOLUTIONS AREN'T EFFECTIVE OR EFFICIENT ENOUGH TO MEET THESE THREATS

**Cybersecurity is labor intensive. Take a look at the steps involved:**

1. Research threats.
2. Identify attacks.
3. Contain breaches.
4. Remediate the damage.
5. Restore systems.
6. Get systems back online.
7. Take steps to ensure the vulnerability is fixed.

This step-by-step task list doesn't even take into account proactive cybersecurity tactics such as creating blacklists and website categorization and the related help desk resources needed to tackle false positives. Traditionally, this has all been done manually with a lot of human interaction. Software simply wasn't smart or adaptable enough to mimic the critical thinking that goes into threat intelligence and response. Risk-averse organizations have elected to err on the side of caution and just lock everything down—*but again, this is expensive and saps user productivity.*

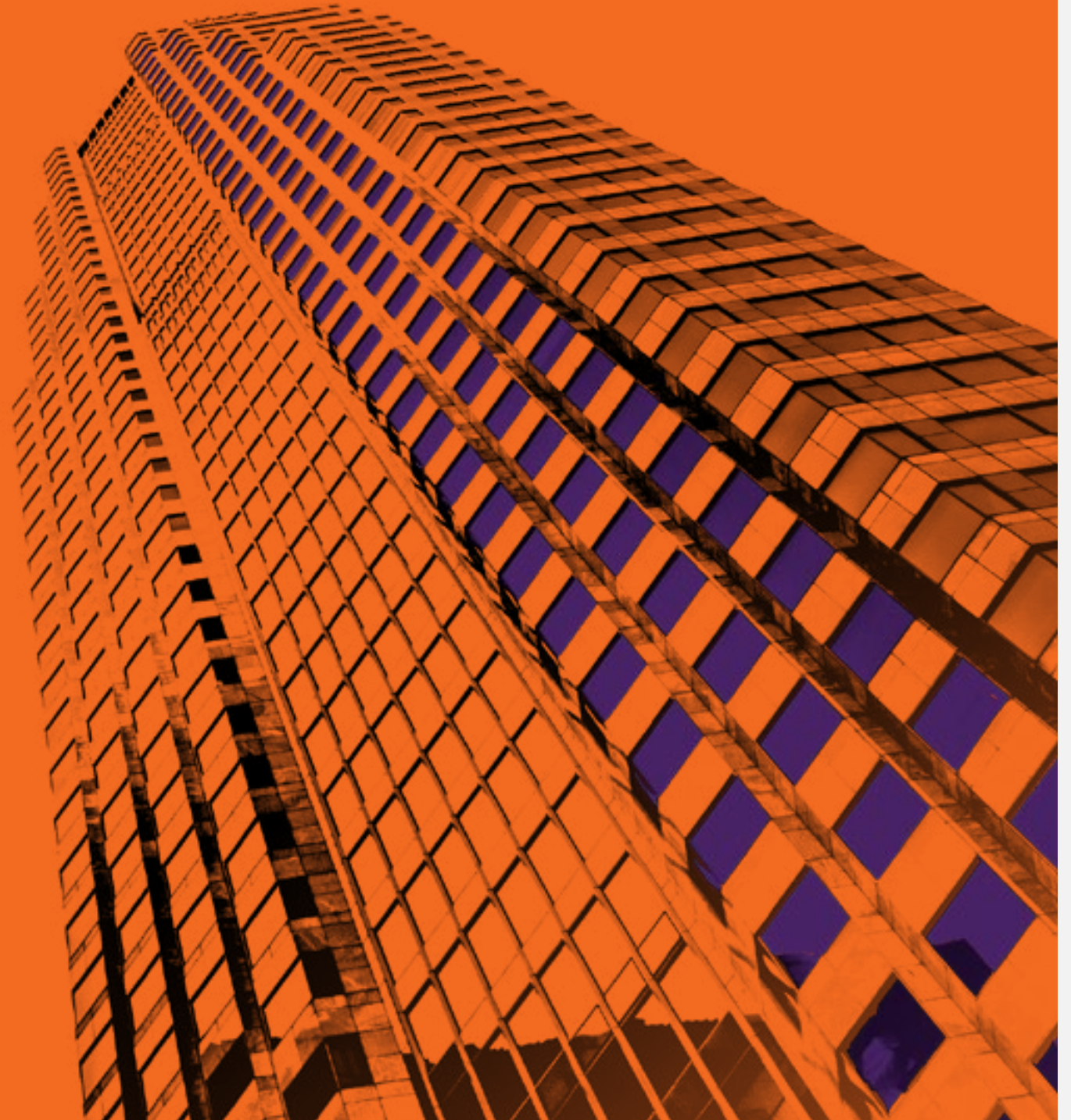
**Organizations are locked into a reactive state.**

Just consider the rapid expansion of new variants of polymorphic malware and crypto ransomware. It's fast-moving, sophisticated, and nearly impossible to stop in real time. To make matters worse, security researchers are forced to rely on data that they can't authenticate while fighting these attacks with manual identification, classification, and pattern matching techniques. The result is that information security teams now find themselves inundated with alerts—many of them false positives—forcing them to chase after an ever-increasing number of security incidents that often seem like wild goose chases. Often, the alerts that are handled are those making the most noise while better disguised (and the most dangerous) threats remain hidden in the haystack.





# **THE SOLUTION: THE POWER AND SCALE OF MACHINE LEARNING**



# THE SOLUTION: THE POWER AND SCALE OF MACHINE LEARNING

Machine learning can help organizations tackle the vast amounts of data collection, data analysis, and prevention needed to protect organizations from today's evolving threat landscape. The right modeling can automate many of the manual tasks that typically bog down today's information security teams. Working in tandem, machine learning and human researchers can keep the organization safe. It's a symbiotic relationship.

**Machine learning allows organizations to comb through the noise to classify information and identify the events that truly need attention:**



Categorize  
websites



Make load  
determinations



Automate  
manual tasks



# WEBROOT: MACHINE LEARNING DNA

Machine learning is part of the Webroot DNA as we have been applying the latest machine learning technology to our advanced cybersecurity solutions for more than ten years. Webroot collects several terabytes of real-time security events daily which is supplemented with petabytes of historic security data to continuously train and update cloud-based models. Because of this history and scalability, URLs, IPs, files, and apps are classified faster and more accurately—serving to both speed up and improve the accuracy of the threat intelligence that powers Webroot's and our partners' solutions. The result is that humans and Webroot machine learning work together to keep Webroot's and our partners' customers safe. Humans are constantly training machines to be more accurate while machines improve human effectiveness and enable scalability.

## Webroot Machine Learning has evolved to its sixth generation:

**1st Gen:** Bayesian

**2nd Gen:** Support Vector Machines (SVM)

**3rd Gen:** Maximum Entropy Discrimination (MED)

**4th Gen:** Active Learning Combined with MED

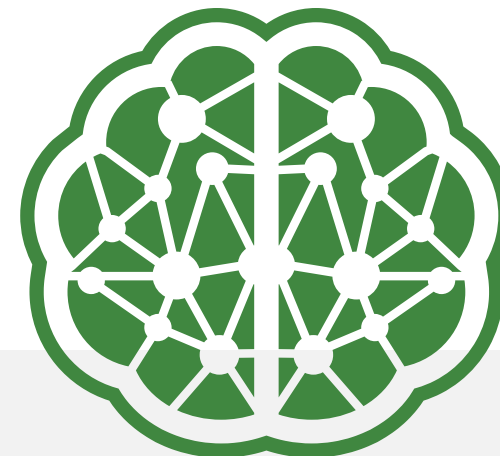
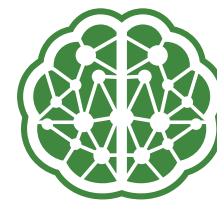
**5th Gen:** Active Feedback Combined with MED

**6th Gen:** Deep Learning

As an industry leader, Webroot laid the groundwork for using machine learning in cybersecurity and is currently using 6th generation machine learning to include deep learning. Webroot deep learning allows us to more accurately and autonomously identify relevant patterns and concepts within continually growing amounts of telemetry from Webroot customers and partners—continuously updating and retraining our models to keep up with the rapidly evolving threat landscape.

Machine learning can be a powerful tool to help solve today's pressing cybersecurity issues. But many companies' machine learning strategy might be categorized as vaporware. It's theory. It's future roadmap stuff. As you have learned, Webroot has a long history of leveraging our machine learning platform to power threat intelligence for our entire portfolio of security solutions. *We're not just talk.*

**Let's take a look at how Webroot machine learning can help organizations navigate this new threat landscape.**



**WEBROOT®**



# THE THREATS ARE GROWING

How to navigate the top five  
security threats.



# 1 **THREAT ACTORS GO AFTER THE MOST VULNERABLE TARGET: PEOPLE**

Email and web-based malware are the most common attack vectors today. Attackers can purchase ready-made phishing templates on the dark web for just a few hundred dollars and unleash a barrage of innocuous-looking emails and landing pages specifically tailored for each individual user. Dozens of emails can be sent with different themes, language, and branding—each one preying upon a different emotion or response—change your password, approve this invoice, check out this funny video, vote for employee of the week. It's a volume business. All they need is a single click to kick start a chain reaction that ends with networks, users or data compromised.

The technology used in phishing attacks is getting more sophisticated all the time. In fact, more phishing sites are using HTTPS to better impersonate legitimate sites, making them difficult to detect. At the same time, the speed at which they can be spun up and taken down make it virtually impossible for organizations to stop them from getting through to users. The only way to prevent an attack is to analyze URLs as users are clicking on a link and make a determination in real time whether to allow or block access.

Webroot machine learning provides the scalability and speed necessary to do this fast enough that it is imperceptible to the user. Powerful models are built to identify how these sites have been created and how they behave, analyzing millions of attributes in microseconds. The IP address, hosting server, how the HTML is constructed beneath the page, the network dialog between client and server—these and thousands

of other factors are used to determine whether a site is legitimate or fraudulent. Models should know, for example, what a real eBay page looks like because it's seen an eBay page hundreds of millions of times. This insight identifies fraudulent phishing sites quickly enough that access can be blocked before a malicious connection is made. The more users surf the web and access legitimate sites, the more a cloud-based model can learn and identify how trusted sites look like and behave.



# 1

## THREAT ACTORS GO AFTER THE MOST VULNERABLE TARGET: PEOPLE

*It can happen to anyone...*

### Fashion House Gets Phished

#### Victim Profile

- Privately held women's apparel co.
- \$4 – 5 million annual revenue
- 20–25 employees

An administrative assistant received an email from the CEO who was on vacation at the time. The email had instructions to wire \$500,000 to a vendor and included the account details. The assistant immediately contacted the finance team, who made the transfer at once.

The email was not from the CEO, and neither the assistant nor the Finance team verified any of the information before making the transfer. When the CEO returned, they contacted the FBI at once, but the funds were never recovered.

**Loss: \$500,000**



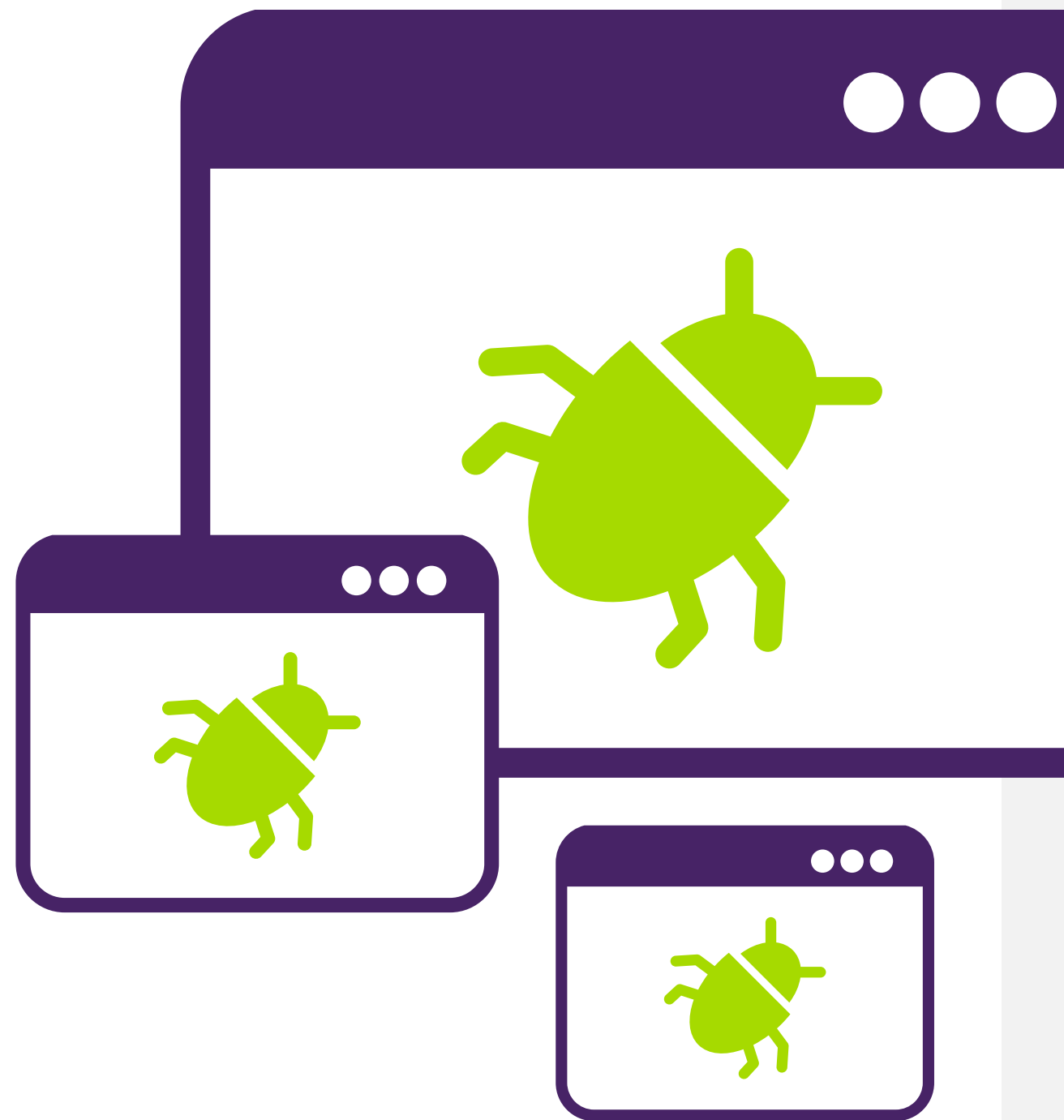
## 2

## ATTACKS ARE INCREASINGLY ON THE RISE AND THEY'RE GETTING MORE DIVERSE

It's no secret that doing business today is dependent on the internet. Users aren't able to do their job unless they can do research online, access email, engage with customers on social media, and log on to cloud-based business systems. The amount of network traffic that this generates is incomprehensible. We're talking about trillions of packets that flow into, out of, and across corporate networks—each one a potential ticking time bomb that can infect systems to grant access to customer data, steal intellectual property, or disrupt operations. It's impossible to monitor and track it all, and if organizations could, it would be cost prohibitive or greatly impede user productivity.

Machine learning automates and scales cybersecurity to meet these growing challenges. Successfully-built models that have been trained properly can learn to quickly and seamlessly flag the needles in a haystack that need further analysis. Researchers can look more closely, take the necessary actions, and then inform the model to catch future iterations of the threat.

Instead of just flagging events that need further analysis, machine learning enabled systems can remediate these threats as well, allowing human researchers to fully focus on higher-level threat intelligence that requires a human touch. Over time, models can be trained to do the work of thousands of threat researchers—combining the specific expertise of thousands of humans in a single, hyper-intelligent entity that can tackle these issues at scale.





# 3

## NEW TECHNOLOGIES CREATE NEW THREAT FRONTIERS

It used to be that organizations could shore up the network perimeter and just monitor traffic as it flowed into and out of their environments. But the move to the cloud has essentially erased any boundaries between “your” network and “their” network. Critical business data can now sit on-premise, in the cloud, or both—making it impossible to simply shore up the castle walls to prevent malicious actors from getting in. Malware can originate inside the corporate network, bypassing traditional perimeter defenses, or sit undetected for weeks, months, or years on a little-used on-prem server before launching an attack. Additionally, most workers are now mobile workers, taking devices home, on travel, and everywhere else they can squeeze in a few minutes of work. The business perimeter is now anywhere an employee goes.

At the same time, the Internet of Things (IoT) has also greatly expanded the threat vector to include disparate machines that are now connected, orchestrating complex processes and tasks. The problem is that IoT security is severely lacking and default passwords are seldom changed. This gives malicious actors a weak link in which to gain access to more hardened critical systems. It’s not hyperbole to think that a vulnerability in the office coffee maker may lead to a major data breach in the CRM system.

Machine learning enables the scalability to monitor all traffic—not just packets moving into and out of the network. Models are built to learn normal behavior of users, servers, systems, and appliances and flag any abnormal behavior that seems suspicious—such as an email server trying to access the CRM system or an industrial system trying to connect with a website in Russia. This can be extremely difficult to do in environments that are controlled by a cloud vendor or in IoT networks where the organization doesn’t fully understand normal versus abnormal behavior. Machine learning can be utilized to better understand how these systems work with each other and flag anything that seems out of the ordinary.





# 4

## THERE'S NO PRECEDENT FOR INCREASINGLY UNIQUE ATTACKS

Spear phishing is a highly-targeted attack that is engineered for one or a handful of people. Malicious actors can do basic Google or LinkedIn research to find a lot of information they can use in a phishing email to make it seem more legitimate. They can find names of co-workers, systems that the user has expertise in, the organizations they belong to—even hobbies, names of spouses, and other personal details. Because the attacks are so unique to the recipient, they are often new and don't look like anything that has been used in a previous attack. This lack of historical data can make it hard to detect—forcing the user to rely on his or her good judgment.

Machine learning doesn't need historical evidence to detect spear phishing attacks. Models can analyze millions of attributes of the links included in an email—from the web server to the underlying HTML of the linked page—to flag a suspicious email or URL as phishing. In addition to this level of threat intelligence, machine learning can even go beyond email and network behavior and analyze the behavior of the sender. Has the CFO ever asked the recipient to wire \$50,000 at the end of the day on Friday? To a human, that would seem suspicious, but a traditional email filter wouldn't even be looking at that. Machine learning is uniquely able to identify and classify the behavior of both technology and people.



## 5 ZERO-DAY ATTACKS EXPLOIT SLOW PATCH SCHEDULES

Zero-day attacks target unknown vulnerabilities and work quickly to exploit them before they are patched. By the time the vulnerability is eventually fixed, it's already too late. Malicious actors have already infiltrated systems. Traditional cybersecurity defenses aren't effective because they are simply unable to react in time.

However, machine learning can render most zero-day attacks ineffective. Instead of monitoring for an attack that it doesn't know is coming, machine learning classification allows organizations to analyze network behavior to detect suspicious activity as a result of a zero-day attack. In actuality, there are very few new attacks each year. Instead, most zero-day attacks are actually old attacks wrapped in something new. Most attacks have a similar goal—they all want access into a specific business system. However, there are only so many ways to do that. It doesn't have to be the same executable or from the same source or author. But they all behave in similar ways—exporting a memory location, exfiltrating data, encrypting files, or building a cryptominer. That's why machine learning models that focus on analyzing and learning behavior and then flagging irregularities are so compelling.



# LET'S SOLVE IT TOGETHER.

Want to discover more? Contact Webroot today to learn more about how Webroot's machine learning powers smarter cybersecurity.

CONTACT US



## WEBROOT®

Webroot was the first to harness the cloud and artificial intelligence to protect businesses and individuals against cyber threats. We provide the number one security solution for managed service providers and

small businesses, who rely on Webroot for endpoint protection, network protection, and security awareness training. Webroot BrightCloud® Threat Intelligence Services are used within our own products and by over 80 market-leading companies like Cisco, F5 Networks, Citrix, Aruba,

Palo Alto Networks, A10 Networks, and more. Leveraging the power of machine learning to protect millions of businesses and individuals, Webroot secures the connected world. Headquartered in Colorado, Webroot operates globally across North America, Europe, and Asia. Discover Smarter Cybersecurity® solutions at [webroot.com](https://webroot.com).

©2018 Webroot Inc. All rights reserved. Webroot, SecureAnywhere, BrightCloud, FlowScape, and Smarter Cybersecurity are trademarks or registered trademarks of Webroot Inc. in the United States and other countries. All other trademarks are properties of their respective owners. WP \_ 020819 \_ US