How Artificial Intelligence Will Secure the 21st Century

Machine Learning and Mathematics Introduce a Brave New World of Predictive Cybersecurity that Rewrites the Rules of Protection

CYLANCE™

Executive Summary

Connectivity defines the world of the 21st century. In just the last ten years alone, social networks, mobile devices, the Internet of Things, and embedded systems have transformed society. You can be virtually anywhere and connect with almost anyone.

For all the achievement and wonder, CXOs and other leaders in the IT space face difficult technological challenges, particularly in the complex world of cybersecurity. Increased threats from constantly evolving attack sophistication and frequency make it difficult to attain a strong security posture. Recent major breaches, such as the one at the United States Office of Personnel Management, have made big headlines, increasing demand for greater security.

The result is that stakeholders are demanding greater levels of cybersecurity. With new cyberthreats released every day, IT executives have more reason than ever to fear compromise, destruction, and manipulation of data.

And it's prompting a massive spending spree. According to Gartner Research, the total market for all security will surpass $100B in 2019.

Against this backdrop, a new technology paradigm based on artificial intelligence (AI) and machine learning is redefining endpoint protection—and possibly security as a whole. By employing prediction and prevention techniques to stop malicious attacks, pre-execution and outside the system architecture, AI is more than just an innovation; it's a complete shift in philosophy and understanding. The results are striking. With efficacy rates at 99%, artificial intelligence and machine learning applied at the endpoint protects at levels never before seen.

The Old World View

> *"We can not solve our problems with the same level of thinking that created them."*
>
> **-Albert Einstein**

For decades, the entire antivirus industry has been built on the mentality of the reactive: systems must suffer an attack before it can be stopped, and protection requires a "sacrificial lamb," or first victim. As a result, legacy antivirus vendors have offered solutions based on that reactionary model. Even the most advanced techniques of signature-based detection, exploit prevention, whitelisting, application controls, and endpoint detection and response, all fall into that victim-first model.

Today, that reactive approach is the industry's greatest weakness.

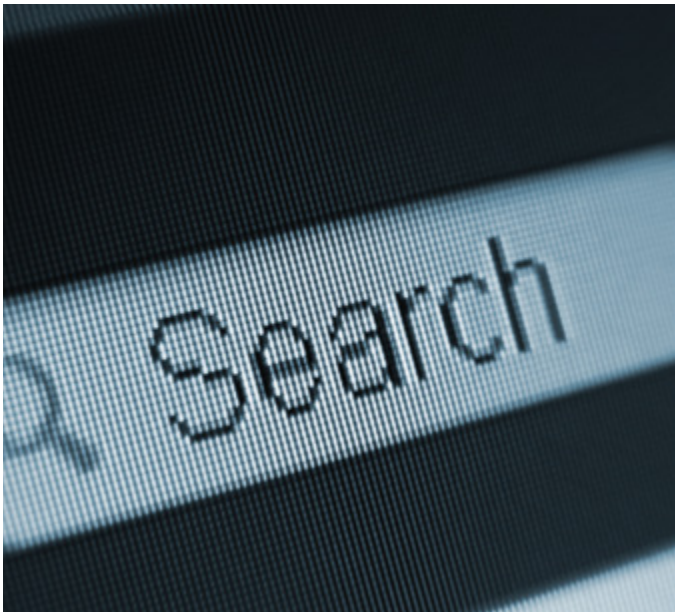
The model of reacting to what has already been seen, experienced, or known is limiting. Because of all the known "unknowns," many companies both large and small have acquiesced when it comes to trying to prevent attacks. The perception is that there are simply too many new techniques and variants of attacks, so people relegate themselves to a response-only mode, and pour precious time and resources into building the fastest response team possible. Remediation is the success measure of the "it's not if, it's when" mentality.

In many ways, the legacy antivirus industry has encouraged this thinking. For all the new software and solutions built over the decades, no uniquely new concepts or ways of thinking have entered the marketplace—until recently.

AI, and its mathematical subset of machine-based learning, are radically changing the old world mode of cybersecurity.

For years, industries such as finance, pharmaceutical, and insurance have successfully applied AI to problem-solving. Modern conveniences benefiting from machine learning and mathematics include online content suggestions, drone piloting, and voice recognition. Perhaps the most used machine learning example is the simple online search. Machine learning algorithms make the keywords and sentences appear as you begin to type your search.

So the question is: why haven't we applied this technology to the world of security?



Previous attempts at machine learning in cybersecurity have failed for a variety of reasons. Often the data samples were lacking, the algorithms too imprecise, or costs too high. Only in the last few years has that changed. Today, the explosion of available data (thanks to IoT and cloud capabilities), increased computing power at lower costs, and development of advanced algorithms, have combined to make applied AI in security a reality. Research firms are taking notice. Gartner recently published the Top 10 Strategic Technology Trends for 2016: Advanced Machine Learning, which acknowledges that machine learning today has made unprecedented progress.

In addition, several market factors have played a role in its utility and adoption.

Why Cybersecurity Matters More in 2016

Major forces at work in the world of security practitioners make endpoint protection even more of a priority today, and give way to the radical approach of AI.

Ransomware

The impact of ransomware is growing. Threat actors encrypt files, rendering data inaccessible until the target pays a hefty ransom. It often relies on social engineering techniques to gain a foothold. The industry can expect to see a dramatic increase in ransomware attacks over the next 12 months because more and more victims come to the same conclusion—the easiest way for getting data back is to pay.

Malware

Modern malware is difficult to trace and often customized to pilfer data. Moreover, a seismic shift from "digital graffiti," to targeted attacks has occurred. Organized crime, nation states, and hacktivists benefit from online communities and resources available on the "dark web." The prevalence of "cybercrime as a service" results in more complex, targeted attacks.

Regulatory Compliance

The vast volume of regional, national, and international rules and regulations have simultaneously prioritized the need for information privacy as well as improved data security. Organizations have more legal obligations and concerns than ever before, and they require technology solutions that meet this increasing legal and regulatory complexity. As an example, The General European Data Protection regulation will increase the responsibility of businesses to handle and protect personal data appropriately. Failure to adhere to the new regulation means that any company suffering a data breach could face millions in fines or up to 4% of their annual global sales.

Protecting Critical Assets

Protecting sensitive data is at an all-time premium in today's modern marketplace. Sensitive data includes private or company information, intellectual property (IP), financial or medical patient information, credit-card data, and other information depending on the business and the industry.

Cybersecurity as a Strategic Initiative

More and more organizations recognize that adopting smart cybersecurity strategies adds to overall business value. IT executives are now being tasked with developing comprehensive plans, processes, and procedures that not only improve security, but also integrate into the corporate mission. In the same way corporate social responsibility became a crucial conversation in modern boardrooms, cybersecurity as a key corporate initiative will also dominate executive discourse.

A New Way of Thinking

> *"If I had asked people what they wanted, they would have said faster horses."*
>
> **- Henry Ford**

With new industry priorities and greater demand for security, what's needed is more than just another tool, technology, solution, or perceived best practice. A radical new way of thinking is needed to redefine the industry.

That's where proactive, predictive, and preventative protection through machine learning comes into play. Machine learning, a subset of AI, uses algorithms to build models that uncover patterns and continually refine them with its learning capabilities. By using machine learning, organizations can make better decisions at a speed and scale that surpass human capabilities. This ability comes from being able to predict based on experiences from the past.

Part of the new proactive method includes focusing on the endpoint, where a majority of attacks take place. Rather than adding more layers of technology to the network, the focus instead should be on the new perimeter of the modern era—the user.

Machine learning offers a comprehensive, granular approach to malware prevention at the equivalent to the DNA-level of code. With human DNA, you have a complex set of instructions which, in blocks (genes), interact with other blocks to create patterns for building a living organism. Machine learning can analyze similarly interrelating blocks of code and file characteristics at a rate and volume that manual analysis by humans cannot begin to match, which means the performance of systems and workflows are not adversely impacted.

One of the marvels of machine learning is that, unlike human analysis, once malware is deconstructed, views of statistically similar blocks of code can be analyzed to identify the presence of malicious code (i.e. "bad genes") without having to execute the file first. AI can determine malicious files through observation, pattern recognition, and predictive analytics. With it, 99% of existing and never before seen malware threats are prevented.

As more organizations understand the value of machine learning versus traditional sandbox, heuristic, or behavioral methods—all of which are reactive, post-execution approaches—they are choosing to invest in it as a pre-execution method of prevention.

To achieve this, machine learning algorithms can be placed on an endpoint host in order to conduct pre-exectution static analysis, which can quickly make a determination if a file is malicious or benign. As opposed to relying on cloud-based analysis techniques, the host can venture off-network and benefit from the same level of protection because the algorithm continues to reside on the host. Unlike techniques such as signature lookups, which require a daily reconnection to the cloud, machine learning algorithms can be fully trained to run off network for months at a time.

Consider why the tide is turning. According to Enterprise Strategy Group's Enterprise Adoption of Next-Generation Endpoint Security: How Enterprises are Evaluating, Testing, and Deploying Security Products (May 2016):

- By 2020, smart machines will be a top-five investment priority for more than 30% of CIOs
- By 2020, CFOs will need to address the valuations derived by smart machine data and "algorithmic business"
- By year-end 2018, 25% of durable goods manufacturers will utilize data generated by smart machines in their customer-facing sales, billing and service workflows
- By year-end 2018, R&D-based end-user approaches to smart machine deployment will be three times more likely to produce business value than IT project-based approaches
- By 2018, more than three million workers globally will be supervised by a "robo-boss"

A New Way of Working: How Machine Learning Works

"Numbers rule the universe..."

- Pythagoras

While the concept of AI sounds more science fiction than science, the precepts are powerfully simple. Machine learning breaks down into four phases:

Collection

The first step in machine learning involves collecting as much data as possible. Virtually hundreds of millions of files are compiled from multiple sources, which include live data feeds, government databases, proprietary repositories, as well as research and scientific surveys that are open source. The ability to gather and store data in the cloud, as well as extract data from mobile, IoT, and embedded systems, facilitates taking data collection to new heights that were not possible even just a few years ago. Gleaning data from all of these sources ensures a relevant sample size that represents the broadest range of file types and authors.

Extraction

The feature extraction process deconstructs a single file into a variety of characteristics that number in the millions. Each characteristic is analyzed against millions of characteristics derived from other files. Thus, millions of records gathered during the collection phase are then individually deconstructed into millions of variables that are transformed into vectors.

Machine learning can be viewed as a kind of self-sustaining network or ecosystem.

To learn, it must observe.
To observe, it must know what to look for.
To know what to look for, it must have previously learned.

While this may seem like circular reasoning to some, it represents the classic puzzle of intelligence, where extracting and parsing techniques build the puzzle by first deconstructing the puzzle pieces.

Extraction plays a role in applying science and engineering capabilities to the process, then scales them by a factor of millions. In the end, it yields the observations that train the AI, and it is through those learned patterns that AI can determine if a new file is benign or malicious.

Learning

After data is collected and features are extracted from each file, the millions of attributes are ready for the learning process. The attributes are converted to numerical values, in the form of vectors, which are used in model training. Dozens of models are created with measurements to ensure the accuracy of prediction, and the testing process itself helps identify ineffective models. Hundreds of millions of files are used to test and validate models. Tested, refined, and ready for action, the final models are loaded for use.

As the files and file attributes go through the learning process, the models develop an understanding of the intention of a sample, which can be used in a predictive fashion to determine the potential risk a new file may pose without having to execute the file itself.

Classification

Building better statistical models allows for highly-tuned classification and clustering. In the end, machine learning relies on precise content categorization. In the case of endpoint protection, the categories are malicious or benign. In addition, classification includes organizing files based on what a file is intended to do. For example, whether a file is intended to perform as a key logger, trojan, etc.

AI can detect subtle statistical connections that to a human may appear innocuous or go unnoticed altogether. This analysis takes milliseconds and is extremely precise because of the breadth of the files and file characteristics analyzed.

The analysis provides a "confidence score" as part of the classification process. The score gives additional insight that can be used to weigh decisions around a single file—such as whether to block, quarantine, monitor or analyze it further.

There is an important distinction between the machine-learning approach and a traditional threat research approach. With machine learning, models are built to determine if a file is malicious or benign, or should be identified as suspicious based on the confidence score. Through the identification of known malicious and benign files—as well as files that may need further investigation—organizations benefit from the powerful capabilities of AI while still having the ability to isolate the few outliers that may require some manual analysis. The enterprise gets the best of both worlds—powerful technology that automates, accelerates, and dramatically improves processes with the option for integrated human expertise when desired. The advantage is that your security staff can move away from responding to thousands of alerts that range in severity to attending the few activities that necessitate their expertise. In addition, you remove from the process the possible bias that impacts legacy antivirus methodologies.

A New Way Forward

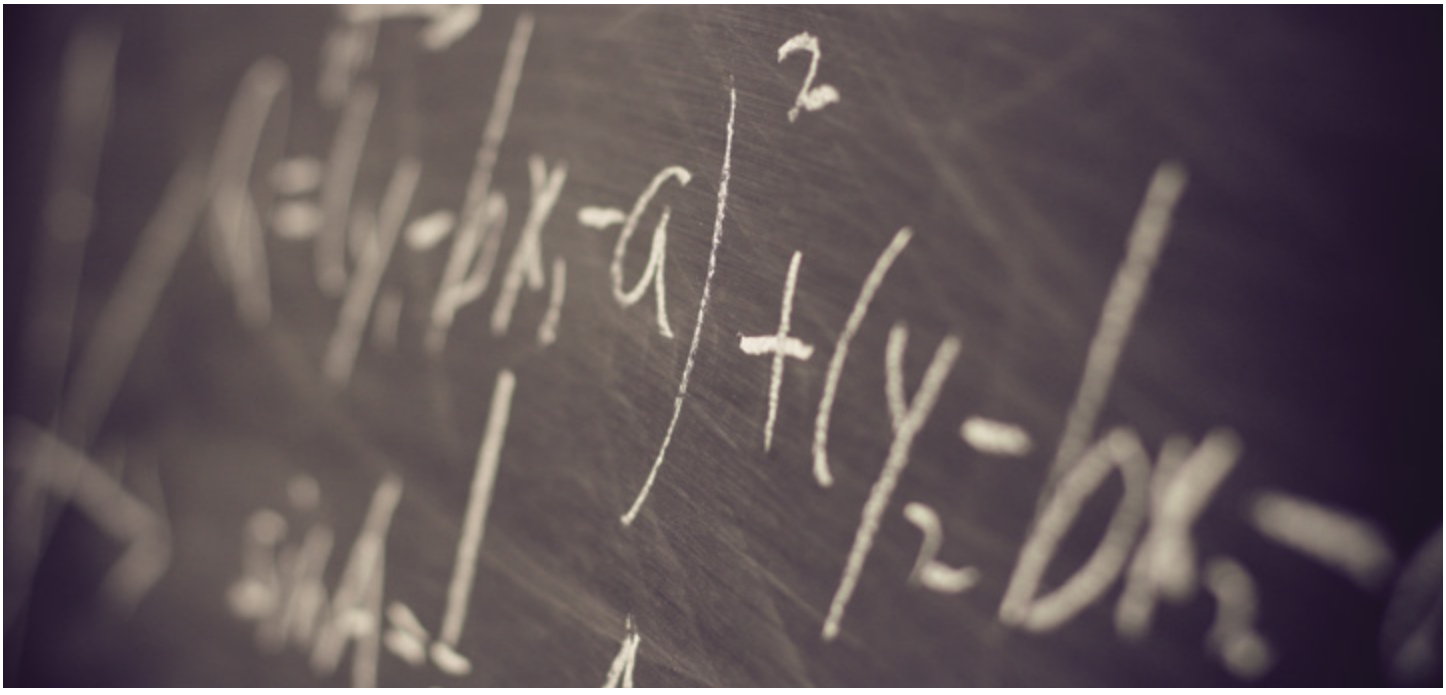
"The value of an idea lies in the using of it."

- Thomas Edison

With AI and machine learning, the future of cybersecurity has arrived. No more legacy antivirus, no more sandboxing, isolation, post-execution response and clean up. Once applied, several core benefits result from a change to your endpoint protection strategies.

Simple

Organizations can protect every endpoint in the enterprise using science-based approaches that require fewer resources and infrastructure. With no signature updates or scan schedules needed, unlike legacy antivirus and endpoint detection solutions, AI capabilities do not require connecting to cloud services to be effective. Administrators can employ threat prevention on a local host without an Internet connection and frequent signature file updates.



Silent

The seemingly endless alerts generated by outmoded techniques significantly drop to give organizations enhanced efficiencies throughout their security operations while also reducing impact to users. In addition, the advantage to employing algorithmic endpoint security is that it uses little memory and percentage of CPU.

When organizations change their cybersecurity approach to pre-execution, they begin to remove layers of technology, resources, and people dedicated to continuous response. As a result, costs are significantly lowered. As companies remove layers and solve issues at the core, they begin to discover ways to consolidate infrastructure.

Scalable

The proactive, predictive, and preventative techniques employed by AI can be applied across platforms, operating systems, file types, and devices. In addition, it can secure Microsoft Windows and Mac OS X and easily integrate into existing SIEM platforms. Solutions are also available in OEM and embedded versions for technology partners, providing flexible solutions for security from threats, including system- and memory-based attacks, malicious documents, zero-day malware, privilege escalations, scripts and potentially unwanted programs.

Conclusion

AI, machine learning, and mathematics provide a powerful combination that changes the endpoint protection paradigm. The previous reactive approaches are being replaced with a proactive new approach, where prevention is the method for protection.

The old paradigm could not fix the core of the problem—so vendors built and supplied layer upon layer of protection that still required a breach or "sacrificial lamb" to discover new threats.

If an organization, no matter its size, can control execution at the endpoint, they can be truly preventative. Thanks to advances in AI, they can predict and protect the endpoint at a rate of success previously unimaginable.

Securing the endpoint with lightning speed and fewer analysts allows IT and SecOps departments to deploy assets within the organization to more useful places, which translates into cost savings.

While vendors in the cybersecurity industry may ask organizations to trust their claims, modern enterprises can take it a step further and trust math and science. These universal principles allow anyone to know the truth about the power of artificial intelligence and machine learning.

To learn more or to request a demo showcasing specific solutions available today, visit www.cylance.com.

+1 (844) CYLANCE
sales@cylance.com
www.cylance.com
18201 Von Karman Avenue, Suite 700, Irvine, CA 92612