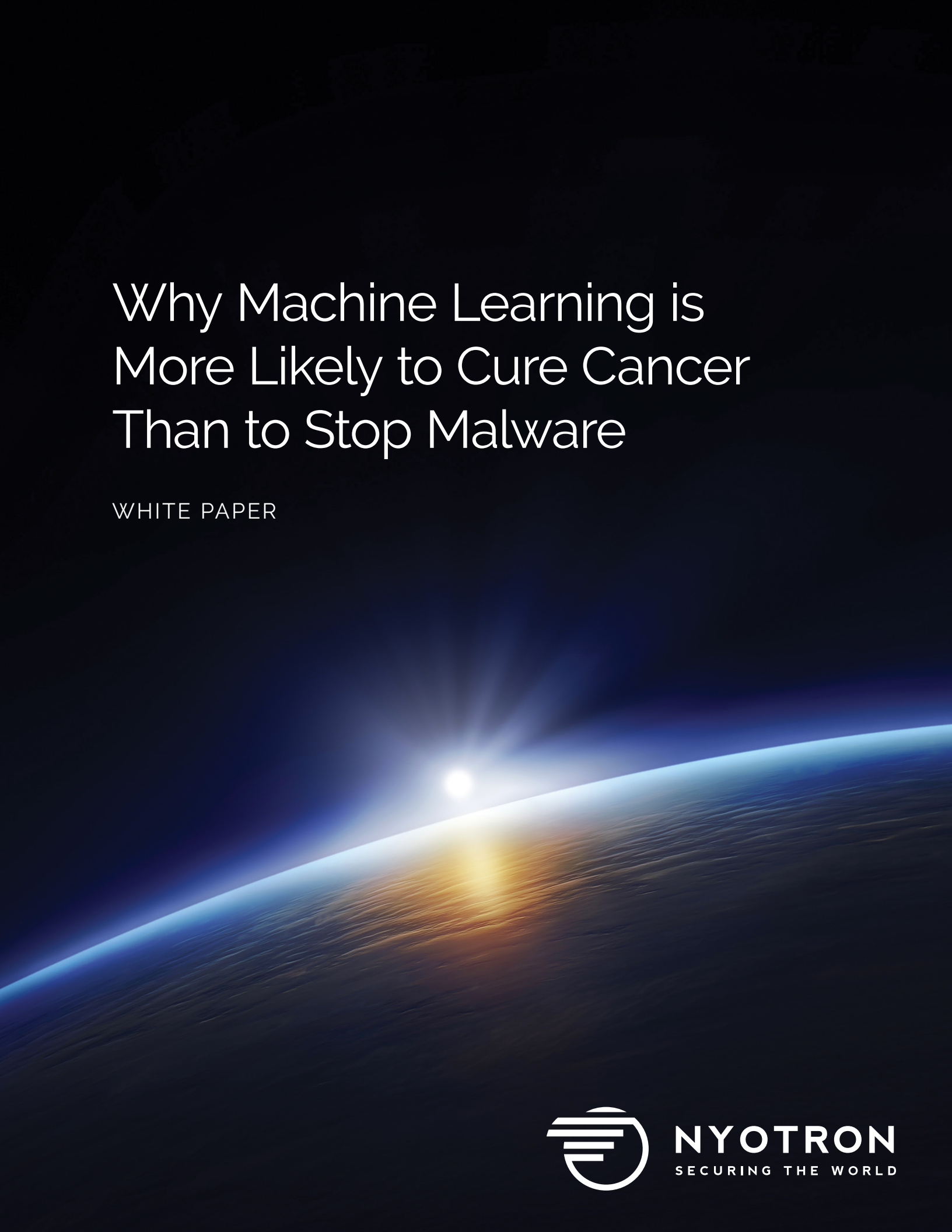
Why Machine Learning is More Likely to Cure Cancer Than to Stop Malware

WHITE PAPER



NYOTRON
SECURING THE WORLD

Introduction

Machine Learning (ML) is based around the idea machines can learn from data. ML techniques have been around for a very long time. In recent years, their use has exploded thanks to advancements in elastic cloud computing and big data. Now, ML is commonly used in fields as diverse as medical research, fraud detection, smart cars, online search and electronic commerce personalization and recommendations, to name just a few.

ML has recently become the shiny new object for security and is the foundational pillar of products such as next-generation antivirus (NGAV) and User and Entity Behavior Analytics (UEBA). While most of these products have promised to be a "silver bullet" against malware, complete protection remains illusive.

In fact, for a number of reasons, ML is more likely to detect and cure cancer than to stop all of today's advanced threats. As far-fetched as this may sound, a detailed look at the distinctions between cancer detection and malware detection quickly demonstrate why this statement is true. Let's have a look at why ML will only get you so far and what you can do about it.

The Past Doesn't Necessarily Predict the Future

ML is being used to detect cancer earlier when it is much easier to treat. This technique can analyze a new image based on past images that are known to show signs of cancer, saving time and improving detection.

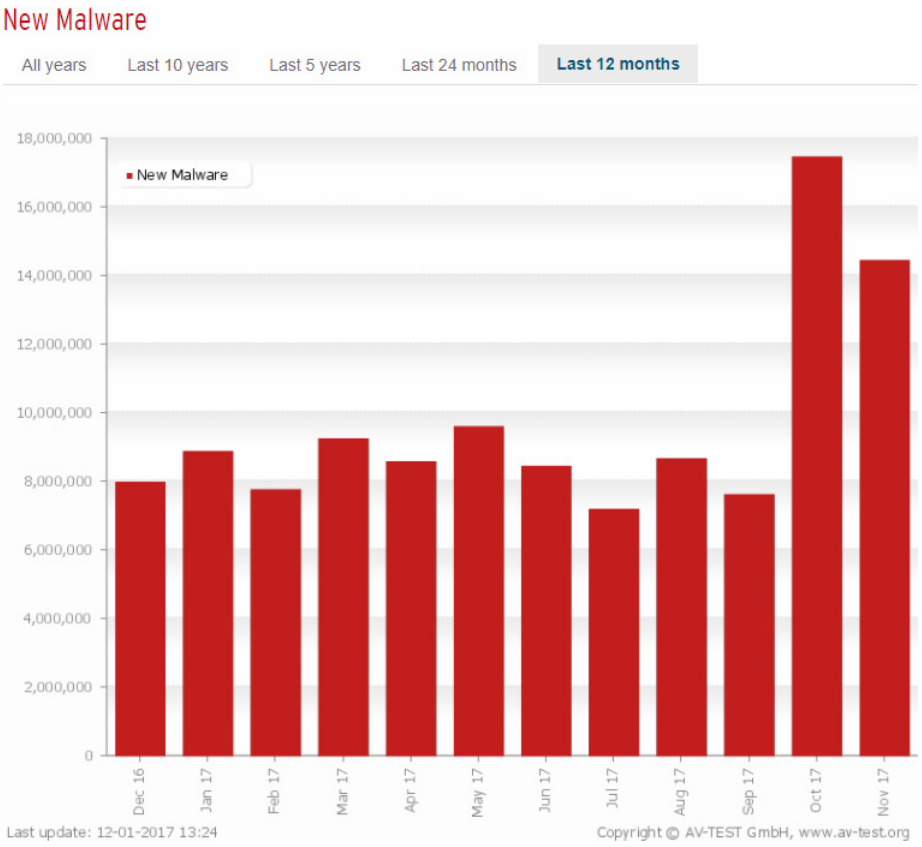
Similarly, security solutions powered by ML are trained on known malware samples in an attempt to detect new malware. However, truly new and previously not seen attack methods and malware techniques appear on a regular basis, making zero-day and advanced persistent threats even harder to detect than cancer.

Moreover, security solutions relying on ML for detection mostly focus on the malware pre-execution step and, thus, attempt to recognize malware by looking at files written to disk. This means that these types of solutions are not very effective against fileless malware that is now being used in nearly 50% of attacks, according to the Verizon 2017 Data Breach Investigations Report.[1]

[1] Verizon 2017 Data Breach Investigations Report, http://www.verizonenterprise.com/resources/reports/rp_DBIR_2017_Report_en_xg.pdf

Good Enough Needs to Be Better

It's always helpful to have a technique like ML that will help improve detection of serious conditions like cancer and serious organizational threats like malware. But, the sheer volume of undetected malware each year can still be overwhelming despite the use of ML-assisted detection. Consider the following example. AV-Test Institute reports that more than 14 million new malware samples were submitted in November 2017 alone.[2] Even a detection rate as high as 99.9 percent would leave more than 14,000 of the new malware samples each year undetected. ML, when applied to malware, needs to be part of a multi-layered security defense system to help fend off attacks that are increasing in number and sophistication.



The Harder You Try, the More You Fail

ML is also being used to eliminate and reduce false positives in cancer diagnosis. For example, ML programs can be used to build a risk assessment model for a patient which is then scored to reduce unnecessary tests and surgeries.

So, is ML equally helpful in reducing false positives for malware detection?

[2] AV-TEST Institute, https://www.av-test.org/en/statistics/malware/

Not usually. Security relying on ML makes the flawed assumption that malware will act the same way that it has in the past. Of course, it rarely does and the endless chore of tweaking ML data models to reflect constantly changing malware behavior is losing the battle to keep up with the never ending flow of new malware targeting your laptops and servers. Fine tuning policies for better detection rates is a futile effort that is only making the problem worse by introducing too many false positives.

To illustrate this point, let's look at what's happening with one type of ML-based malware detection technology, Endpoint Detection and Response (EDR), that is infamous for producing lots of false positives. According to Gartner,

> "Time-stretched and resource-constrained CISOs and Security units simply don't have the bandwidth to proactively wade through reams of EDR data hunting for threats and figuring out how to respond to them."[3]

Nothing Will Keep the Bad Guys Out

Security layers act like gates. The problem with gates is that as soon as attackers bypass them, they can do whatever they want with an organization's data. There is no stopping the bad guys. Nevertheless, vendors of ML-based security solutions proclaim that their gate is the one that can stop the next cyberattack. A better gate. A smarter gate. A NEXT GENERATION GATE. Yet, despite implementing these gates, organization still aren't safe.

You Can't Always Be Connected

ML applied to cancer detection as well as other areas typically relies on big data processing performed either in the cloud or on on-premise servers. Its requirement for network/Internet connectivity isn't an issue because it doesn't negatively impact results.

However, this isn't the case in endpoint security. Network connectivity requirements to block attacks leaves a huge void in coverage since endpoints may be unprotected from new, previously unknown malware when not connected to their security service/server. ML-based security solutions require a connection at some point in time to on-prem or cloud infrastructure for activities, such as behavioral model updates, threat intelligence feeds as well as sandbox detonation of previously unseen files.

[3] Gartner Blog Network, Booming $500 Million EDR market faces stiff challenges, https://blogs.gartner.com/avivah-litan/2017/01/12/booming-500-million-edr-market-faces-stiff-challenges/

The Bad Guys are Turning Your Own Weapon Against You

Just like security vendors can train their machine learning models on malware samples to detect them, malware writers can "train" or tune their malware to avoid detection. Attackers can also poison the data that machine learning models use. Because machine learning algorithms use massive amounts of data, it can be difficult to weed out these efforts. Criminals might also steal a machine learning model and reverse engineering it to better understand how to create samples that avoid detection.

You Can Do Even Better

Although ML is of great help to improve malware detection, one technique like ML can't do everything. You can significantly mitigate your security risks by adding an ***OS-Centric Positive Security Model*** to your endpoint threat defenses, whether these are ML-driven or traditional antivirus solutions.

An OS-Centric Positive Security Model is a game changing force in security. Instead of chasing after an infinite number of malware samples and attack vectors, a different approach to malware detection is to focus on the finite ways they can be used. A blueprint of finite legitimate system behavior can be used to understand all the normative ways that may lead to damage. This makes it possible to distinguish between "good" and "bad" actions—regardless of the threat type, attack vector or its origin—to detect and prevent malicious activity.

An OS-Centric Positive Security Model has a number of additional advantages over ML techniques:

- Offers better detection rates and fewer false positives since what constitutes legitimate system behavior is more readily understood than millions of unknown malware samples and attack vectors.
- Seamlessly compliments anti-malware solutions to offer a second line of defense
- Replaces easily bypassable gates with continuous event sequence monitoring
- Requires no cloud or server connectivity to block attacks
- Provides a pre-defined map of all permitted behavior, thus, avoiding the complexities of fine-tuning data models

[4] Gartner's 2017 AI Development Strategies Survey

Summary

To wrap up, ML isn't security nirvana. It's unlikely to solve all malware detection challenges anytime soon for a number of reasons:

- The past doesn't predict the future
- Good enough needs to be better
- Nothing will keep the bad guys out
- The harder you try the more you fail
- You can't always be connected
- The bad guys are turning your own weapon against you

We've shared just a few details of a new approach to detecting malware, what we call an OS-Centric Positive Security Model. Please visit ***www.nyotron.com*** or email us at ***info@nyotron.com*** for more information about this innovation in endpoint security.

About Nyotron

Nyotron is a privately held cybersecurity company that has developed a disruptive Threat-Agnostic Defense™ technology to cope with the biggest challenge of today's digital era—the unknown threat. PARANOID is designed to prevent targeted and advanced national-level cyber-attacks on high-profile enterprises, and it does so without any previous knowledge about the threat or its methodologies. Based on a unique last-line-of-defense approach, the company's technology is designed to protect enterprise data and critical assets by mitigating threats that are able to outsmart all security layers. Nyotron's customer base includes all major industries.

NYOTRON
SECURING THE WORLD

2880 Lakeside Drive Suite 237
Santa Clara, CA 95054
+1 (408) 780-0750
www.nyotron.com