Outsmarting Malware

Why Machine Learning Is Critical to Cybersecurity

Table of Contents

Executive Summary

New strains of malware are constantly threatening businesses and creating angst for IT. As cyber risks grow in both volume and sophistication, the tools used to find and eradicate them have to get smarter and scale better, too.

This paper discusses a modern approach to cybersecurity that uses adaptable machine learning algorithms combined with several anti-malware technologies to find and foil advanced threats. It compares how these components work and their respective attributes to help you determine best practices for the threat prevention component of your organization's security strategy.

Introduction: The Ever-Changing Face of Malware

Malware writers are always a step ahead of traditional security solutions, creating threats that behave differently from system to system, day to day, and year to year. For example, some infections now disguise or even partially encrypt themselves so they don't match known signatures in the malware databases.

Newer malware types include adware, botnet loaders, information stealers, and tech support scams. Another type, ransomware, is growing in popularity in part because of the recent rise of cryptocurrency—digital money that allows anonymous Internet buying and selling. Ransomware encrypts your data and prevents you from accessing it until you pay a digital fee within a specified time period.

How well is our industry combating threats like these? A recent industry study found that fewer than one in three enterprises rate their cyberattack protection as highly effective. Clearly, a stronger set of security systems and processes is needed.

Why We Need a New Approach to Threat Prevention

More new software programs are being written than ever before, and traditional antivirus systems have difficulty detecting brand-new threats that target them. This issue only promises to get worse. To defend against it, a newer approach—one that is automated and can discover the new risks—is needed.

The earliest threat prevention solutions were antivirus systems based on signature matching. A monitoring system would search for a match to a known malicious software signature. If a match was detected, the system would alert an IT expert and possibly quarantine or block the traffic.

These systems turned out to be too narrow in their discovery criteria to have a very high success rate at catching malicious software; their effectiveness was mostly limited to known and documented malware. For example, the companies that have suffered high-profile breaches over the past several years were all running up-to-date antivirus software.

As time went on, and Internet use became more widespread, malware writing became more popular. Antivirus defense was no longer enough to stop the barrage of new threats. These early systems evolved to monitor more complex signatures and then added specific rules to supplement or replace those signatures. Rules combine program attributes and logic to identify the features that indicate a malicious file. Because the rules don't apply to 100 percent of all situations, exceptions to the rules had to be created and entered into the system.

Antivirus Software: Necessary But Not Enough

Antivirus software scans computer files as they are opened, closed, and e-mailed in an effort to detect and remediate malware infections. Most antivirus software—today more accurately called "anti-malware" because it scans for a variety of threat types—uses two primary techniques: searching a signature/rules database; and monitoring for suspicious behavior.

- **Signature/rules database search.** Signature-based antivirus tools look for known patterns that have been identified and stored in a malware database. If a piece of code in the file matches a virus identified in the database, the software usually takes one of three actions: it deletes the file entirely; it quarantines the file so it can't spread; or it attempts to repair the file by removing the virus.
- **Program behavior monitoring.** Some antivirus tools also monitor the behavior of all computer programs in search of anything that deviates from "normal" activity. If one program tries to write data to an executable program, for example, this action is flagged as suspicious behavior. The user is alerted to the action and asked what to do. This function provides a measure of protection against brand-new viruses not yet logged in malware databases. However, it also generates a large number of "false positive" alerts—sending warnings of activity that is unusual but turns out to be legitimate.

Because of their complexity and vulnerability to skewed weighting, rules-based anti-malware systems are quite susceptible to missing a threat. They also tend to generate a very large volume of false positives that must be investigated, squandering the time of valuable security personnel.

Malware researchers “weighted” some of these rules to indicate a higher level of suspicion and establish an overall level of risk for a sample. However, these weighted rules were typically either too aggressive or not aggressive enough.

Due to the rules’ increasing complexity and imperfectly calculated weighting, such anti-malware systems have been susceptible to missing threats and generating false-positive alerts about benign software.

In a networked business environment, the greater the traffic volumes, the greater the number of alerts. Many turn out to be dead ends, resulting in highly valued security personnel wasting time chasing irrelevant alerts and overlooking genuine threats. As a result, security personnel began to regularly ignore alerts. While this is understandable, it defeats the purpose of the security system.

There is so much data being generated now—researchers estimate that the digital world doubles in size every two years—there aren’t enough people to investigate all the alerts. Rules-based systems are complicated to maintain and, at the end of the day, simply don’t work well enough to justify the effort required to keep them current.

Enter Machine Learning

The effects of Moore’s Law and Kryder’s Law now allow for large-scale malware storage and analysis. In the current phase of anti-malware evolution, some vendors are starting to overlay machine learning onto the rules-based systems described. In systems that rely on weighted rules, the weights can now be better optimized by machine learning techniques than by human intuition or pen-and-paper statistics. Additionally, many novel techniques are being developed that take advantage of recent breakthroughs in artificial intelligence, increased computational power, and large consolidated datasets.

Juniper Networks has capitalized on all of these developments. Building a machine learning pipeline from the ground up rather than relying on human-generated rules allows us to learn directly from sample data. It also allows machine learning to be integrated across all threat prevention products in the security suite so they can all benefit from classification-optimized algorithms.

Machine learning integrated into security tools from their inception stands a stronger chance of finding and thwarting new attacks before they can cause significant harm. Such systems continually and dynamically learn what’s “normal” in software structure, software behavior, and network traffic patterns and usage. Millions of variables and data points can be analyzed at once to identify abnormal behavior that could indicate an attack.

Juniper advocates combining traditional signature-based detection with machine learning to catch both known threats and unknown, still undocumented malware. To this end, Juniper has created a machine learning system that uses both static and dynamic analysis of malware signatures to identify new threats. These tools are being deployed in the cloud so that machine learning models can be quickly updated, retained, and applied to the ever-changing threat landscape. As a result, the system can quickly detect and stop new threats before they are identified or analyzed by the industry at large.

How it Works

The machine learning system is fed a large number of signals extracted from various sources ranging from network information to binary structure to runtime behavior. The system learns to weight these signals and map them to a severity/danger potential, which can culminate in a decision to quarantine a system that is likely compromised. The nature and weighting of the used signals can be determined automatically by the system itself.

One advantage of this approach is that the more data that is fed into the system, the better it can distinguish malicious programs from benign ones. Rules that uniquely identify each malware family no longer have to be manually written. Instead, the system identifies specific useful signals generated from program structure, behavior on the system, and behavior on the network, and uses the collected intelligence to separate benign software from malware.

As mentioned, Juniper advocates applying machine learning to security tools beyond antivirus software to include the following:

- **Static Analysis.** This component of threat mitigation analyzes software without executing code. It looks at sample data and how code is structured, then tries to figure out if it is bad, good, or something in between. This is a challenging problem because modern obfuscation techniques can make it nearly impossible to predict the behavior of a sample without running it. However, machine learning techniques can still ascertain the trustworthiness of a sample, its similarity to known samples, and even its degree of inscrutability.

- **Sandboxing.** A sandbox emulates the operating system and runs the executable. In a rule-based system, if the sample doesn't do anything malicious in a short time frame, the system might assume it is benign. If the sample behaves abnormally, such as encrypting or exfiltrating user data, it is likely malware. However, the sample might not do anything malicious but still be labeled "bad" because its behavior (or lack of behavior!) can be statistically correlated with malicious activity. That's the advantage of machine learning over explicit algorithms.

  Used alone, sandboxes have many challenges and may require manual researcher oversight. One challenge is that many Indicators of Compromise (IOCs) are ambiguous. For example, legitimate software will often do things that seem anomalous—such as using undocumented APIs to deliver a desired user experience—but in the end are not malicious. This is why it's crucial to use machine learning to interpret the results.

  Another challenge is malware evasion tactics. Some malware writers will program their malware to lie dormant for a given length of time to fool the sandbox. For example, if there's a typical time span for analyzing binaries for malicious activity, the software can be programmed to wait until after the analysis is complete to execute. These kinds of evasion tactics are designed to fool behavioral signatures and are in a perpetual battle with the antivirus system's heuristic rules. With machine learning, it's possible to identify the evasion itself, even if the sample does not ultimately perform all of its malicious actions. In other words, the very act of inaction/sleeping is a signal in itself. Millions of such signals are gathered, weighted implicitly by the machine learning system, and combined with all collected data to help make an overall decision.

The Juniper Approach

Juniper Networks Sky Advanced Threat Prevention cloud-based solution detects malware and mitigates threats. Unlike many other security systems, which started out simplistic and evolved over time, Sky ATP was purpose-built to take full advantage of modern and innovative machine learning techniques.

Sky ATP includes the information and identifiers that traditional threat prevention tools use but, in addition, takes advantage of ambiguous structural and behavioral properties of potential malware to determine maliciousness.

Early in the Sky ATP analysis pipeline, each new sample is run against a suite of antivirus engines, which is a fast and efficient way to catch and filter out known threats and their close variants. Removing these known threats from the analysis pipeline as early as possible reduces the load on the more computationally expensive parts of the pipeline, which include static analysis engines and full sandbox detonation.

Traffic is fed to the cloud from customers' Juniper Networks SRX Series Services Gateways. This way, changes required to adapt to the current threat landscape are made centrally, and customers do not have to change out their firewalls.

Sky ATP contains some differentiators that put Juniper ahead of the competition:

- Sky ATP uses machine learning across all detection techniques.
- The system uses a number of innovative techniques to lure malware into revealing itself, which measurably increases our detection rate.
- Sky ATP also detects software communicating to unusual servers and evaluates that activity.
- A full networking hardware portfolio—routers, switches, and firewalls—gives Juniper a much richer set of data and behavior, far beyond what is available to vendors who only offer standalone security appliances.

Cyber Security Methods: A Summary

Anti-Malware Method	Description	Strengths	Limitations
Antivirus software (traditional)	Quickly finds and counters known malware using a signature database and rule-based heuristics	• Finds known threats quickly • Finds new malware with heuristics-based methods • Low false-positive rate	• There is a lower chance of finding new ("zero-day") threats than other approaches.
Sandboxing (part of dynamic analysis)	Allows you to observe code in action without it affecting production systems	• Can apply techniques that trick malware into activating and self-identifying • Unnecessary to unpack software • Complete oversight over sample behavior • Rich data ranging from syscalls to network analysis • Can monitor network activity	• It is the slowest method. • Anomalous behavior is not always bad. For example, some software writers optimize video game performance by writing performance-critical parts in assembly, an unorthodox approach that might look suspicious to sandbox solutions. • It is a needle-in-a-haystack solution with many things going on in a live operating system. • There is the potential for self-identification and self-infection.
Static analysis	Analyzes application software without executing code	• Generally works faster than antivirus • Can find new unknown threats • Finds known threats quickly • More difficult to exploit • Not susceptible to active behavior-based evasion techniques	• Can be overcome by specially designed packers. • Potentially easier to generate false alarms if not retrained frequently because structure and nature of software changes dramatically over time.
Machine learning	Applies dynamic, adaptive learning to the many variables and behavior detected by AV systems, static and dynamic analyses, and reputation, identifying new threats before they do damage	• Automatically identifies previously unknown malware • Combines the capabilities of a number of cybersecurity approaches • Scales the knowledge of skilled human analysts to large data sizes • Reduces false positives, leaving fewer and more apparent decisions for human analysts to make	• Requires specialized skills and expertise to implement. • System must be continuously updated, tuned, and monitored to ensure maximum effectiveness. • Not all vendors use machine learning across the same cybersecurity tools. • Effectiveness is only as good as the data used to train the system; in other words, "garbage in, garbage out."

Conclusion

While machine learning alone isn't a magic bullet, it fundamentally changes the security equation by dramatically improving the accuracy of malware detection and risk classification.

Of course, these techniques don't completely obviate the need for human decision making, but machine learning can perform the bulk of manual labor, scaling the knowledge of skilled human analysts to large data sizes and handling the complexity beyond human capabilities. In this way, it can be combined with other security methods to let each of them scale appropriately while keeping misleading alerts to a minimum.

Going forward, the scale of data and complexity of analysis are only going to increase. Machine learning is the only tool available that can tame attacks at the massive scale that has descended upon us.

About Juniper Networks

Juniper Networks challenges the status quo with products, solutions and services that transform the economics of networking. Our team co-innovates with customers and partners to deliver automated, scalable and secure networks with agility, performance and value. Additional information can be found at Juniper Networks or connect with Juniper on Twitter and Facebook.

Corporate and Sales Headquarters
Juniper Networks, Inc.
1133 Innovation Way
Sunnyvale, CA 94089 USA
Phone: 888.JUNIPER (888.586.4737)
or +1.408.745.2000
Fax: +1.408.745.2100
www.juniper.net

APAC and EMEA Headquarters
Juniper Networks International B.V.
Boeing Avenue 240
1119 PZ Schiphol-Rijk
Amsterdam, The Netherlands
Phone: +31.0.207.125.700
Fax: +31.0.207.125.701