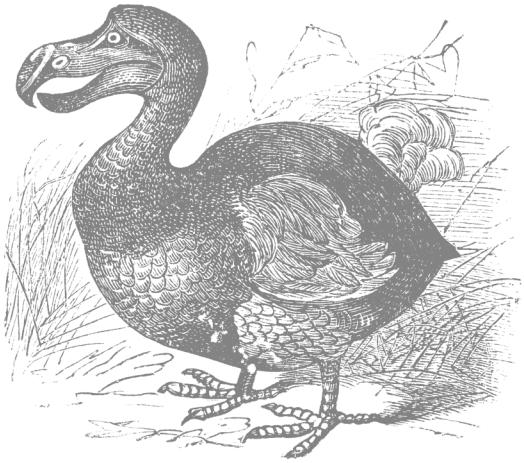
CYLANCE

Applying Data Science to Advanced Threats

An Overview of the Cylance Infinity Platform

September 13, 2013



The Problem

The cyber security industry is now over 30 years old. And just like people, with each passing decade, we realize that what worked for us in our 20s, simply won't work for us now or going forward. In fact, carrying forward the mindset and behaviors of those first 20 years exposes us to countless problems in health and long term solvency. We learn that to survive in the world we must adapt and evolve to a higher form of existence. The antiquated and archaic practices of our past limit our visibility into the future in detecting, and thereby avoiding, maliciousness. Consequently they have given rise to a freight train sized hole of opportunity for the cyber criminals, nation states and cyber miscreants that wish to exploit our blindspots in the cyber world.

Blacklisting (and Signatures) Can Be Compromised

Blacklisting technologies rely almost 100% on signature based techniques for detecting bad files have been at the heart of our industry since the beginning when we had only rare outbreaks like Michelangelo, Stoned and the Morris Worm. The grossly unfortunate fact is that they remain the predominant form of detection (and thereby prevention) in the market today. Signature based approaches to security served us well then when the number of bad objects (files, network traffic, and vulnerabilities) was small and the techniques to alter those files to bypass detection were non-existent or at least non-trivial.

Today however, countless techniques exist to avoid these once stalwart protection technologies, including packers, mutation engines, obfuscators, encryption and virtualization bypass techniques.

Within milliseconds, a once easily detected malicious file can be altered to be completely invisible to even today's best detection technologies while remaining functionally identical to its original maliciousness. This allows the bad guys to easily bypass security infrastructure that once detected them with ease.

The sheer numbers of files submitted to security vendors today for analysis (over 100k daily) is so overwhelming that most vendors simply cannot handle the volume. Their methods and manpower become easily avalanched over. The scale of the problem outnumbers the industry's capacity for maintenance. As a result we have rampant miss rates.

Whitelisting Can Be Compromised

Whitelisting technologies developed in response to what the Blacklisting world is victim to: low detection rates. In other words, ***blacklisting alone detects only 5-10% of malicious files out there.*** The reason whitelisting was so promising for so long was that it effectively did the opposite of blacklisting: rather than stopping everything known bad (which is large and hard to do), whitelisting only allowed to run those files which are known good (which is much smaller and presumably easier to do). This technique has been applied to security through identification of permissible URLs and files that are known (or perceived) to be clean and safe. But these solutions have some fundamental problems as well.

The first challenge with solutions that rely heavily on whitelisting is that one must simply "trust" what the vendor (or your operations staff) has designated as "good". We have seen this model fall down time and again with security and software vendors who have their development environments compromised and their private signing certificates stolen (e.g. Adobe, Bit9 and Opera Software). When these attacks occurred it allowed the thief to sign their own malicious files as if they came from the "trusted" vendor. And because whitelisting solutions rely so heavily on this "trust" model, it allows the bad guys to easily bypass the technology.

Trust Can Be Compromised

As a consequence to the identified gaps of blacklisting and whitelisting, numerous technologies have crept up to fill in the gaps of signature technology including host intrusion protection systems (HIPS), heuristics, behavioral, and both hardware and software sandboxing. But all of these techniques have two core weaknesses: **1)** foundational signature elements, and **2)** reliance on "trust".

Technologies such as HIPS, heuristics and behavioral engines remain at their core, signature based. They rely on "knowing" what is bad and creating a signature for that "badness". Even sandboxing technologies which claim no signatures are involved to auto-detonate captured files and binaries, still rely on signatures to enable alerting and blocking the next time it sees it.

For these technologies to know if something is good or bad, they must map them to a list of known good or bad behaviors which can take minutes, hours, or days using manual verification. Even then, the attack has already happened and the detonation may not discern the maliciousness of the malware.

Can we simply "trust" our vendors to show us what is "good"?

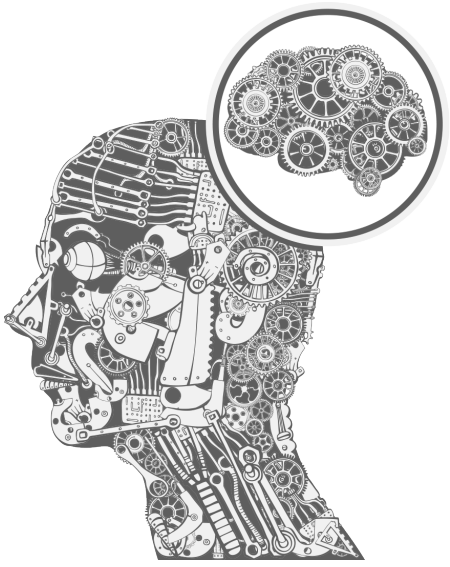
The Need to Evolve Security

Bad guys have the advantage in more resources and time to outwit the various detection schemes of security vendors. Additionally, many security models (like signatures) require the engagement of a human. Human involvement is fallible and limited in scale to the speed and sophistication of advanced threats.

Can we simply "trust" our security vendors to show us what is "bad"?

We as an industry must evolve from this outlived model to a new and ever-evolving technique; one that abandons signatures and blind trust; one that relies on a mathematical, algorithmic and scientific approach to better effectiveness and measurable accuracy.

In short, we must evolve to "Trust the Math" and science of Cylance's Infinity.



Introducing Cylance Infinity

Infinity is a fundamental and epic shift from traditional security methods of detecting good and bad. It is a highly intelligent, machine-learning, data analysis platform.

As battle tested security industry veterans, we know that the previous approaches can never cope with the volume and variety of advanced threats. So we designed Infinity to make intelligent decisions without relying on signatures. It does this by taking a predictive and actuarial approach to data on a network to determine good from bad.

This model exists in many other industries. Insurance companies use actuarial science to determine the likelihood of a risk event for the insured person at a surprisingly high rate of accuracy. This concept relies on advanced models of likely outcomes based on a variety of factors. For a standard insurance policy, they may consider twenty to thirty facts to determine the most likely outcome and charge appropriately. Infinity uses tens of thousands of measured facts harnessed across millions of objects to make its decisions, in near real-time.

Infinity, at its heart, is a massively scalable data processing system capable of generating highly efficient mathematical models for any number of problems.

Cylance uses these models applied to 'big data' to solve very hard security problems with highly accurate results at exceptionally rapid rates. It's done by applying data science and machine learning on a massive scale. Coupled with world class subject matter experts, cyber security is able to leap ahead of threats.

While Infinity is problem agnostic, correctly designing solutions to hard problems takes time, knowledge and effort. The Cylance Infinity Labs team has focused all of their efforts on detecting advanced threats, in near real-time, correctly, without signatures.

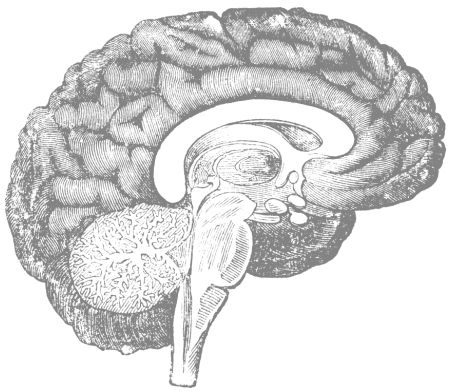
This problem is one that has long plagued the whole Internet. The existing "solutions" involving humans, trust models or signatures have proven vastly incapable of solving this problem, resulting in massive infections, data loss, and a hostile environment for business, consumers, and the internet at large.

What is Machine Learning?

Machine Learning (ML) is a formal branch of Artificial Intelligence and Computational Learning Theory that focuses on building computer systems that can learn from data and make decisions about subsequent data.

In 1950, Alan Turing first proposed the question, "Can computers think?" However, rather than teaching a computer to "think" in a general sense, the science of machine learning is about creating a system to computationally do what humans (as thinking entities) do in specific contexts.

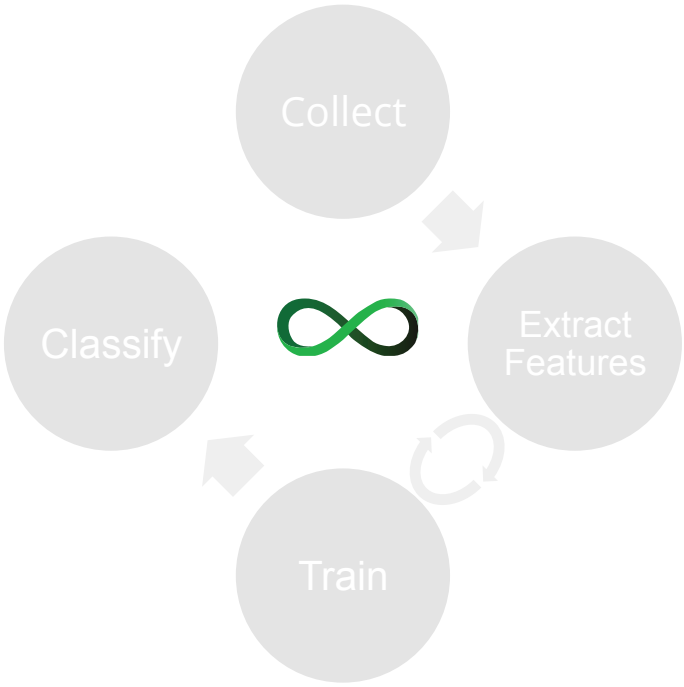
Machine Learning (ML) and big data analytics go hand-in-hand so ML focuses on prediction, based on properties learned from earlier data. This is how Infinity identifies malicious versus safe or legitimate files. Data mining focuses on the discovery of previously unknown properties of data, so those properties can be used in future ML decisions. ***This means Infinity learns on a continual basis, even as attacker methodologies change over time!***

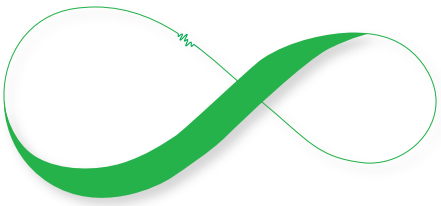


How it Works

Infinity collects data, trains and learns from the data, and calculates likely outcomes based on what it sees. It's constantly getting smarter from environmental feedback and a constant stream of new data from all around the world.

To achieve its magic, Infinity performs the following steps. First it **COLLECTS** vast amounts of data from every conceivable source. Second, Infinity **EXTRACTS FEATURES** that we have defined to be uniquely atomic characteristics of the file depending on its type (.exe, .dll, .com, .pdf, .java, .doc, .xls, .ppt, etc.). Third, Infinity constantly adjusts to the real-time threatscape and **TRAINS** the machine learning system for better decisions. Finally, for each query to Infinity, we **CLASSIFY** the data as good or bad.





Infinity – The Rubber Meets the Road

Infinity be used to supercharge decision making at endpoints, and woven tightly into existing security systems via a variety of integration options. It is cloud enabled (but not cloud dependent) to support advanced detection on a global scale in limited form factor environments, or can operate autonomously while still achieving a stunning rate of protection.

The breadth of deployment options helps to solve several fundamental problem points on a modern network.

CylanceV and CylanceV Local

CylanceV is a REST SSL Application Programing Interface integration to Infinity's intelligent cyber security decision making. Through the API and specially developed utilities, IT departments executing incident response and forensics can take the tedium out of tracking down malware and determining what is truly bad.

CylanceV enables a starting point for forensic analysis and timely remediation through an automated and highly efficient approach.

Tying other security tools like SIEM, Log analysis, host and network monitoring, HIPS/NIDS and investigation tools including anti-virus, anti-malware and forensics, into CylanceV provides contextual intelligence for more accurate and effective malware identification.

The CylanceV API allows utilities to be developed in most popular frameworks (.NET, Python, etc.) and invoked through HTTPS using tools such as CURL or WGET in order to make the data segmentation easier and more efficient.

CylanceV *Local* is an on-premise version of CylanceV that allows for use in restricted and sensitive environments.

Integrating 3rd party functionality, like Python scripts, Splunk, C# to Infinity quickly determines what is safe and what is a threat, making smart security smarter. Together, they reduce the total number of prospective compromised machines to something manageable.

Infinity On the Endpoint

Cylance***PROTECT*** is our host based security solution built on Infinity technology. It leverages algorithmic science to greatly increase the speed and accuracy of host protection without reliance on signatures, heuristics or behavior modeling. It offers a real-time protection layer on the endpoint that can make decisions about the nature of malware independent of connecting to Infinity and at a stunningly low performance impact. ***PROTECT*** offers a powerful front line of defense, whether your assets are behind your corporate firewall or in a coffee shop. Its extensive management capabilities easily blend the pervasive protection into your existing security workflow.

Summary

With Infinity, we can definitively determine good or bad file objects milliseconds, with extraordinarily high detection accuracy and extremely low false positive rates. Because the system is self-collecting, self-training, and self-learning, we always stay ahead of the changes and unknowns attempted by the bad guys. With such a mathematical approach, we may change the game of security... **forever.**

About Cylance

Cylance is a global cyber security products and services company headquartered in Irvine, California. Its founders, Stuart McClure and Ryan Permeh, know that today's network and operations infrastructure is inadequately protected by flawed security.

Stuart is a leading authority in information security and lead-author of "Hacking Exposed: Network Security Secrets and Solutions". Stuart launched the vulnerability assessment leader Foundstone, Inc. and served as Global CTO at McAfee as well as EVP/GM of the Security Management Business Unit.

Ryan is a leading expert in development of security technologies who, with Stuart, built TRACE, McAfee's elite threat team, and unique detection technologies. Both have witnessed the security industry's evolution firsthand over the past 25 years and know that the security infrastructure for today and tomorrow's threats is fundamentally broken.

Cylance is driven by an impressive team of veteran Security executives, board of directors and advisors, and deeply talented security professionals to achieve a simple mission: Solve the world's most difficult security problems.

Cylance, Inc.

+1 (877) 973-3336
sales@cylance.com
www.cylance.com

West Coast Office:

46 Discovery, #200
Irvine, CA 92618
USA

East Coast Office:

11710 Plaza America Drive, # 2000
Reston, VA 20190
USA