The Essential Guide to the MITRE ATT&CK Evaluation Round 3

This e-book provides a comparative look at how vendors performed across various measures, with guidance on how to explore the results further. We include key descriptions of MITRE's testing methodology, the tools MITRE provides to help visualize and compare results, and considerations for analysis to help you assess for yourself which vendor best fits your organization's endpoint security needs.

paloalto® NETWORKS | CORTEX® XDR BY PALO ALTO NETWORKS

Introduction

To prevent and protect against today's sophisticated and craftiest cyberthreats, one must first understand the threat actor behind the keyboard. To combat modern advanced persistent threats (APTs) where adversarial behavior is changing continually, vendors need an objective format in which to test their capabilities against the tactics, techniques, and procedures (TTPs) being used across the threat landscape.

The MITRE ATT&CK® evaluations provide the opportunity to do just that, effectively analyzing the ability of leading endpoint detection and response (EDR) solutions as well as extended detection and response (XDR) vendors and their products to detect and defend against real-world attack sequences.

For the third year running, Palo Alto Networks has emerged as one of the top-performing vendors in the MITRE ATT&CK Evaluations, delivering 100% threat protection and more than 97% detection visibility—with zero Configuration Changes.[1]

At a high level, Cortex XDR achieved the following against the TTPs used by Carbanak and FIN7:

- 100% blocking of attacks in the protection evaluation on both Windows® and Linux endpoints.
- 97% visibility of attack techniques.
- The best detection rates of any solution that also got a perfect protection score.
- 86% analytics detection, defined by MITRE as detections that provide additional context beyond telemetry, based on the attack techniques used in the evaluation.
- 80% of detections having an associated technique-level detection, the highest type of detection awarded in this evaluation.
- The highest overall combined detection and protection rate in the evaluation.

Cortex XDR blocked

100%

of attacks in the protection evaluation on both Windows and Linux endpoints

CORTEX XDR BY PALO ALTO NETWORKS

Evaluation Overview

For Round 3 of the MITRE ATT&CK Evaluations, MITRE tested a larger field of vendors compared to the previous two years, providing further evidence of the importance of third-party evaluations in the marketplace for objective guidance around choosing security solutions.

These evaluations provide assessments for participating vendors to identify areas for improvement, including updating prevention, detection, and response rules that inform security policies. While this exercise does not provide overall comparison scores or ranking, it provides a vendor-agnostic summary of the various methodologies employed by security practitioners for identifying and preventing sophisticated attack campaigns.

Testing 29 vendor participants, covering 20 separate test steps with 174 sub-tests on both Windows and Linux operating systems, the evaluations pitted each vendor against the TTPs leveraged by the Carbanak and FIN7 threat groups.

What's Different This Year?

In Round 3, about half of the vendors participated in a separate part of the evaluation that was focused on protection capabilities that spanned Windows and Linux, with 10 steps where products were tested to see if they could actively block attacks. Given our track record for excellent threat prevention and our extensive tooling for Linux endpoints, we opted to participate in the protection tests.

In the evaluation results, Cortex® XDR™ **blocked all attacks across both Linux and Windows while providing the highest detection rate and quality of detections** of any vendor to do so (see figure 3). By highlighting endpoint protection as a standalone evaluation, MITRE underscores the value of looking beyond traditional detection capabilities, giving credence to the importance of combining an endpoint protection platform (EPP) with EDR for a more complete endpoint security solution.

Cortex XDR achieved 97% visibility
of attack techniques—
the best detection rates
of any solution that also
got a perfect protection score

CORTEX XDR BY PALO ALTO NETWORKS

MITRE's Approach

Focused on articulating how detections occur rather than assigning scores to vendor capabilities, MITRE categorizes each detection and capture.[2] Detections are then organized according to each technique. Techniques may have more than one detection if the capability detects the technique in different ways, and detections they observe are included in the results.

While MITRE makes every effort to capture different detections, vendor capabilities may be able to detect procedures in ways that MITRE did not capture. For a detection to be included for a given technique, it must apply to that technique specifically. For example, just because a detection applies to one technique in a step or sub-step, that does not mean it applies to all techniques of that step. For proof of detection in each category, MITRE requires that the proof be provided to them, but they may not include all detection details in public results, particularly when those details are sensitive.

To determine the appropriate category for a detection, MITRE reviews the screenshot(s) provided, notes taken during the evaluation, results of follow-up questions to the vendor, and vendor feedback on draft results. They also independently test procedures in a separate lab environment as well as review open source tool detections and forensic artifacts. This testing informs what is considered to be a detection for each technique.

After performing detection categorizations, MITRE calibrates the categories across all vendors to look for discrepancies and ensure categories are applied consistently. The decision of what category to apply is ultimately based on human analysis and is therefore subject to discretion and biases inherent in all human analysis, although efforts are made to hedge against these biases by structuring analysis as described herein.

Did You Know?

In 2013, MITRE ATT&CK got its start from MITRE's Fort Meade Experiment (FMX) where researchers impersonated adversarial groups' tactics and techniques?

Using MITRE to Help Evaluate EDR Solutions

For organizations who are reviewing EDR solutions and vendors, the MITRE results present a comparison of the various levels of security efficacy by participating vendors, all aligned around a common lexicon to ensure parity and continuity across the evaluation.

So, how can the MITRE ATT&CK evaluations help inform a defensive strategy for solution providers like us? At Palo Alto Networks, participating in the MITRE ATT&CK evaluations provides us the opportunity to be tested by a neutral, unbiased third party, leveraging current sophisticated attack sequences that yield constructive insights into how we can build more effective detection and prevention solutions.

In using the modern attack TTPs from groups such as Carbanak and emulating the attack scenario in a controlled environment—the MITRE Engenuity-provided cyber range—solution providers can assess their performance and determine areas for improvement. The resulting performance data can provide insights into solutions or product modifications and give guidance for fine-tuning any steps that may have underperformed.

About the Adversary

Modern bank robberies don't get much slicker than the blatant financial sector APT-style attack campaigns perpetrated by Carbanak, which netted around US$1 billion from hitting nearly 100 financial institutions worldwide from 2013 to 2018. They are sometimes referred to as FIN7 due to using the namesake Carbanak malware, yet FIN7 and Carbanak appear to be two different groups and are therefore tracked separately.

Carbanak malware netted a

US$1 billion

haul in financial sector attacks involving threat actors from Russia, Ukraine, Europe, and China

Primarily using spear phishing emails with malicious attachments sent to bank personnel, Carbanak was able to infect systems, steal credentials, harvest intelligence (such as intercepting bank clerk screens), and mimic staff in order to steal funds in a variety of ways, including:

- Transferring money to fraudsters' accounts via online banking.
- Making e-payments, transferring money to accounts in the US and China.
- Inflating account balances and transferring the differences via fraudulent transactions.
- Controlling ATMs to dispense cash at predetermined times for henchmen to collect.

At a Glance: MITRE's Carbanak/FIN7 Emulation

- 2 complete scenarios (one per adversary)
- 20 attack phases
- 174 sub-steps with 70 unique techniques
- Protection evaluation (10 steps)

To view the in-scope techniques for the Carbanak+FIN7 evaluation in the ATT&CK Navigator, MITRE provides the layer file available here.

Credential Access	Discovery	Lateral Movement	Collection
Account Manipulation	Account Discovery	AppleScript	Audio Capture
Bash History	Application Window Discovery	Application Deployment Software	Automated Collection
Brute Force	Browser Bookmark Discovery	Component Object Model and Distributed COM	Clipboard Data
Credential Dumping	Domain Trust Discovery	Exploitation of Remote Services	Data from Information Repositories
Credentials from Web Browsers	File and Directory Discovery	Internal Spearphishing	Data from Local System
Credentials in Files	Network Service Scanning	Logon Scripts	Data from Network Shared Drive
Credentials in Registry	Network Share Discovery	Pass the Hash	Data from Removable Media
Exploitation for Credential Access	Network Sniffing	Pass the Ticket	Data Staged
Forced Authentication	Password Policy Discovery	Remote Desktop Protocol	Email Collection
Hooking	Peripheral Device Discovery	Remote File Copy	Input Capture
Input Capture	Permission Groups Discovery	Remote Services	Man in the Browser
Input Prompt	Process Discovery	Replication Through Removable Media	Screen Capture

Carbanak FIN7 Carbanak+FIN7

MITRE Round 3 Methodology

The Environment

The evaluations were performed in Microsoft Azure Cloud. Each vendor was provided with two identical environments consisting of eight hosts each on which to install their client software. These two environments were used for the detection-only and protection tests, respectively. The vendors also had the option of installing server software onto a virtual machine (VM) already in the environment or importing a VM if necessary. By default, the Azure VMs were Standard B4MS, each with four vCPUs and 16GB memory. Each vendor had full and complete administrative access to the hosts instantiated for them.

VPN access enabled connectivity to the environment, and passwords were shared via out-of-band methods. There was one VPN server per environment and vendors then used RDP or SSH elsewhere within the environment. Hosts were reachable only within the VPN. They did not have public IP addresses assigned to them via Azure, but they were able to access the internet.

Continued on next page >

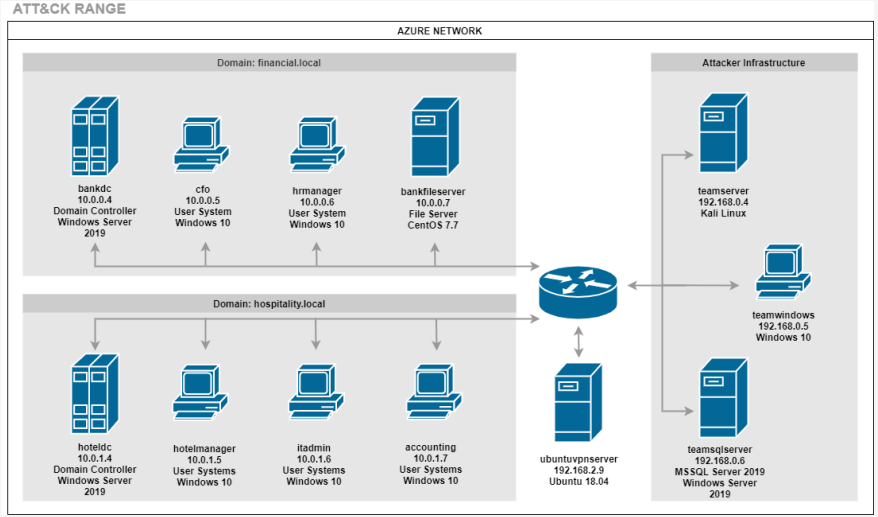

Figure 1: ATT&CK Range—Azure network

Carbanak/FIN7 Evaluation Environment

Target Hosts:

- Windows Server 2019
- Windows 10
- CentOS 7.7

While each participant may use their own unique terminology and approach to detect and protect adversary behavior, MITRE abstracts and summarizes the respective data into two main categories to discuss the products in similar terms: "Main" and "Modifier."

In relation to the amount of context provided to the user, each detection or protection receives one main category designation, although one or more modifier category designation can be provided to help describe the event in more detail as an option.

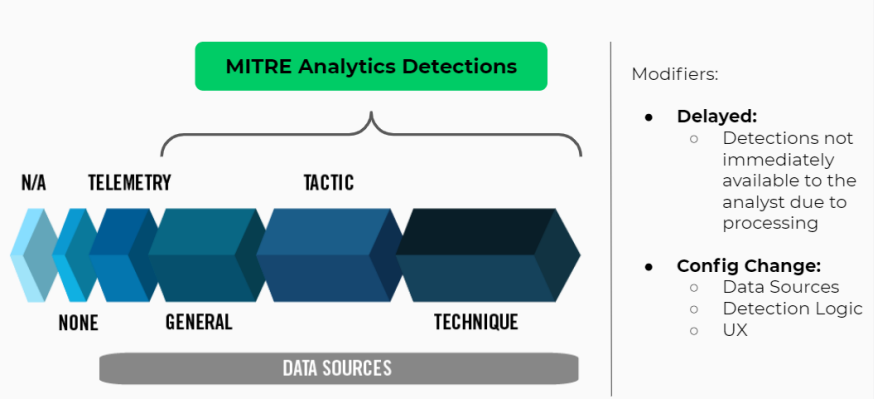

Figure 2: Carbanak/FIN7 detection categories

For the Carbanak+FIN7 evaluation, there are *six main detection categories* representing the amount of context provided to the analyst, and *three main protection categories*.

Continued on next page >

Detection Categories

Not Applicable: Vendor did not have visibility on the system under test. The vendor must state before the evaluation what systems they did not deploy a sensor on to enable Not Applicable to be in scope for relevant steps.

None: No data was made available within the capability related to the behavior under test that satisfies the assigned detection criteria. There are no modifiers, notes, or screenshots included with a None.

Telemetry: Minimally processed data collected by the capability showing that event(s) occurred specific to the behavior under test that satisfies the assigned detection criteria. Evidence must show definitively that behavior occurred and be related to the execution mechanism (did happen vs may have happened). This data must be visible natively within the tool and can include data retrieved from the endpoint.

General: Processed data specifying that malicious/abnormal event(s) occurred, with relation to the behavior under test. No or limited details are provided as to why the action was performed (tactic), or details for how the action was performed (technique).

Tactic: Processed data specifying ATT&CK Tactic or equivalent level of enrichment to the data collected by the capability. Gives the analyst information on the potential intent of the activity or helps answer the question "why this would be done". To qualify as a detection, there must be more than a label on the event identifying the ATT&CK Tactic, and it must clearly connect a tactic-level description with the technique under-test.

Continued on next page >

Technique: Processed data specifying ATT&CK Technique, Sub-Technique, or equivalent level of enrichment to the data collected by the capability. Gives the analyst information on how the action was performed or helps answer the question "what was done" (i.e., Accessibility Features or Credential Dumping). To qualify as a detection, there must be more than a label on the event identifying the ATT&CK Technique ID (TID), and it must clearly connect a technique-level description with the technique under-test.

Protection Categories

Protection categories were used to identify whether a protection was encountered in the adversary emulation, and whether a user prompt was required to confirm the blocking activity. Categories are subject to change, based on lessons learned from the evaluation.

Not Applicable: Vendor did not deploy protection capabilities on the system under test. The vendor must state before the evaluation what systems they did not deploy a sensor on to enable Not Applicable to be in scope for relevant steps.

None: The technique under test was not blocked and/or the technique was unsuccessful and there is no evidence provided to the user that the capability blocked the activity.

Blocked: The technique under test was blocked and the user was explicitly informed that the capability blocked the activity.

Continued on next page >

Modifier Detection Types

MITRE differentiates between types of detection to provide more context around the capabilities a vendor offers in a way that allows end users to weigh, score, or rank the types of detection against their needs, enabling end users of the results to decide what is most beneficial.

Configuration Change: The configuration of the capability was changed since the start of the evaluation. This may be done to show additional data can be collected and/or processed. The Configuration Change modifier may be applied with additional modifiers describing the nature of the change, to include:

- **Data Sources** – Changes made to collect new information by the sensor.
- **Detection Logic** – Changes made to data processing logic.
- **UX** – Changes related to the display of data that was already collected but not visible to the user.

Delayed: The detection is not immediately available to the analyst due to additional processing unavailable due to some factor that slows or defers its presentation to the user, for example subsequent or additional processing produces a detection for the activity. The Delayed category is not applied for normal automated data ingestion and routine processing taking minimal time for data to appear to the user, nor is it applied due to range or connectivity issues that are unrelated to the capability itself. The Delayed modifier will always be applied with modifiers describing more detail about the nature of the delay.

Cortex XDR vs. Carbanak+FIN7: Our Results

Focused on analyzing how detections occur rather than assigning scores to vendor capabilities, MITRE categorizes each detection and capture, and then organizes detections according to each attack technique. Techniques may have more than one detection if a security solution detects a technique in different ways. All observed detections are included in the evaluation results.

MITRE combined the attack techniques detected by telemetry (meaning little processing was required to detect the technique) and those detections that required analytic processing to determine "visibility" to arrive at the overall detection rate of the 174 attack techniques the vendors were tested on.

The Cortex XDR Difference: The Data Doesn't Lie

As the industry's first XDR platform, Cortex XDR integrates endpoint, network, cloud, and third-party data to stop sophisticated attacks. As evidenced by our leading results in the MITRE ATT&CK Evaluations for three years running, Cortex XDR achieved high performance in protection, detection, and visibility—the pillars for a holistic and best-in-class endpoint security solution.

Cortex XDR provides increased detection fidelity with behavioral analytics and machine learning. It collects and stitches together a broad set of data, including logs from Cortex XDR endpoints, Next-Generation Firewalls, Prisma® Access, identity providers, and much more. Cortex XDR builds a profile of expected user behavior to pinpoint unusual behavior indicative of attack. Behavioral analytics applies machine learning and statistical analysis to rich data to uncover attacker tactics and techniques with fewer false positives than traditional detection rules.

Cortex XDR: Stellar Evaluation Results, Three Years Running

- **2018:** Broadest coverage across attack techniques
- **2019:** Unsurpassed overall attack technique coverage
- **2020:** Best combined detection and prevention

CORTEX XDR BY PALO ALTO NETWORKS

Because Cortex XDR combines protection, analytics detection, and visibility, anomalous behavior is precisely identified, expediting the triage process as well as reducing dwell time and subsequent lateral movement within a network.

Protection Is the Foundation

Cortex XDR not only blocked all attacks in the first-ever MITRE ATT&CK protection tests; it also integrated log data from Palo Alto Networks Next-Generation Firewalls to increase detection fidelity. Because protection means prevention, the adversary was unable to execute the attack, resulting in zero dwell time. Furthermore, stopping the threat reduces alert fatigue as follow-up steps do not occur, disrupting the sequence of the attack lifecycle.

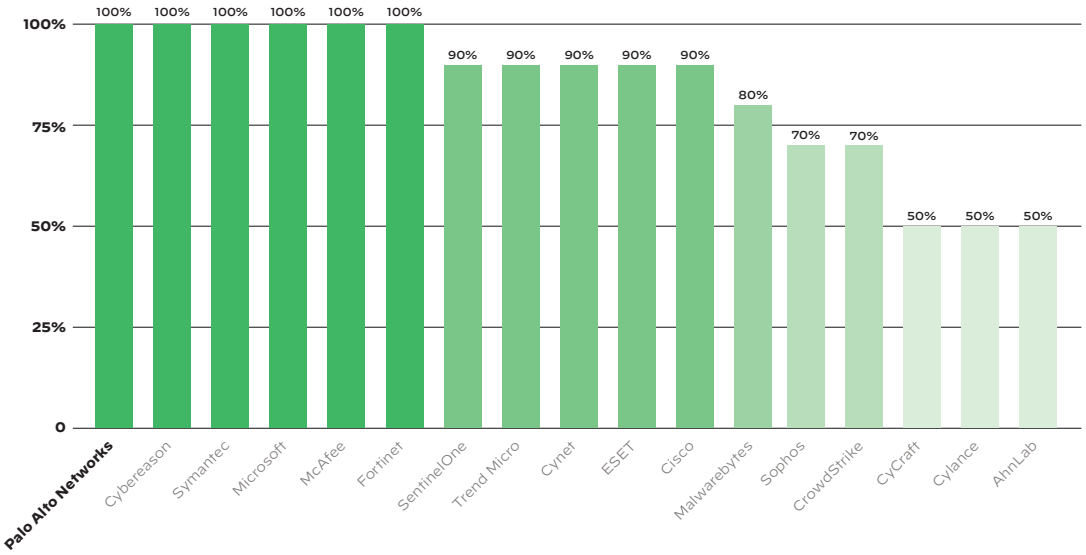

Figure 3: Cortex XDR blocked 100% of attacks in the protection phase against both Linux and Windows

CORTEX XDR BY PALO ALTO NETWORKS

Cortex XDR: Best Overall Performance

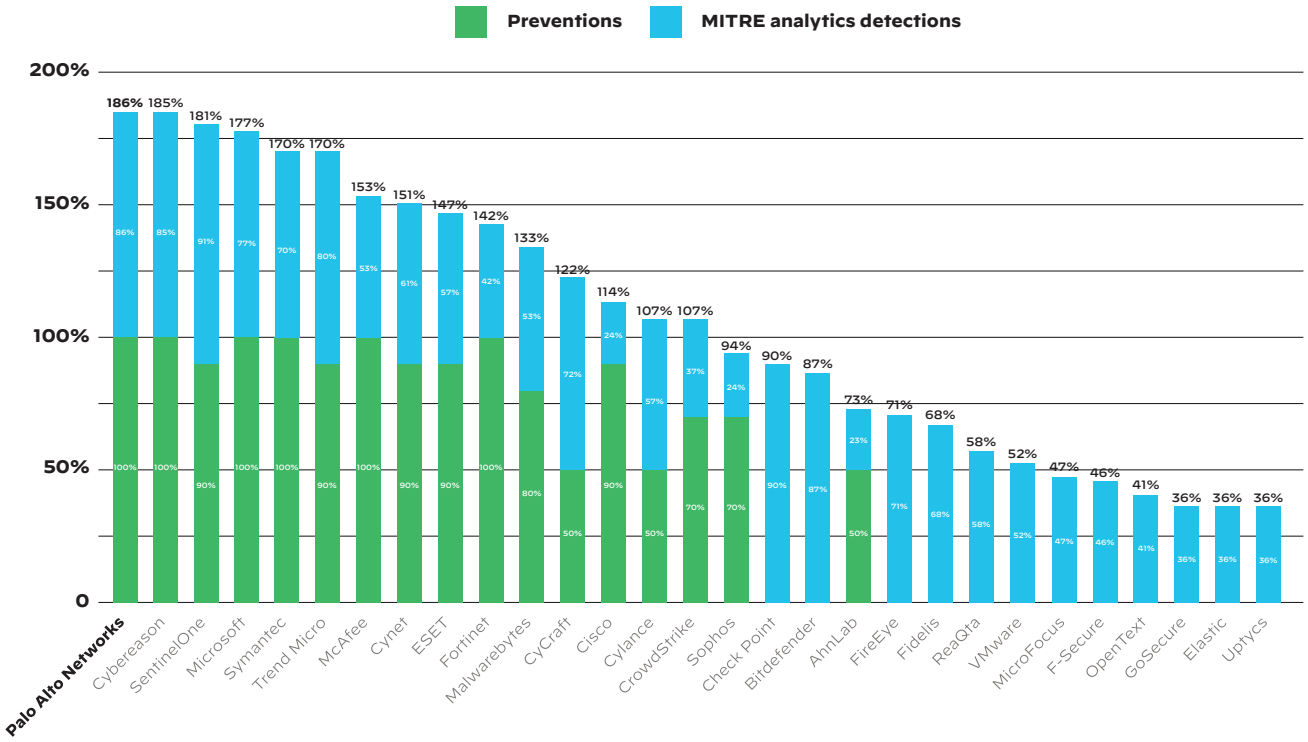

Figure 4: Best combined protection and analytics detection

CORTEX XDR BY PALO ALTO NETWORKS

Strong protection or prevention is critical in an EDR solution, significantly reducing the impact to security analysts to help free up time for investigation and threat hunting. Great detection provides visibility into the attack sequence, delivering the right analytics to sift through and pinpoint anomalous activity that warrants further investigation. Visibility is the foundation of prevention and detection, but visibility alone often amounts to just noise. When analytics is leveraged to stitch together and correlate telemetry from multiple sources, attack campaigns become clearer.

About Configuration Changes

MITRE allows for solution providers to have a "do-over" if a step in the evaluation did not produce the desired detection. These do-overs are called "Configuration Changes." This allows security vendors to improve their detection against a technique they did not detect with their initial configuration. Therefore, a Configuration Change is simply a detection that was made possible because a change was made in order to garner a better result. MITRE provides this opportunity for vendors so they will have the chance to validate how changes to the solution may improve security efficacy.

In the real world, when an attacker is not caught executing a step in their attack chain, they don't give you a second chance with a new configuration to catch them in the act. For that reason, we feel it is best to exclude detections (see figure 5) directly resulting from a Configuration Change when comparing results.

Examples of Configuration Changes[3] include:

- A new rule is created, a pre-existing rule enabled, or sensitivities (e.g., block lists) changed to successfully trigger during a retest. These would be labeled with the modifier "Configuration Change-Detection Logic."
- Data showing account creation is collected on the backend but not displayed to the end user by default. The vendor changes a backend setting to allow telemetry on account creation to be displayed in the user interface, so a detection of Telemetry and "Configuration Change-UX" would be given for the Create Account technique.

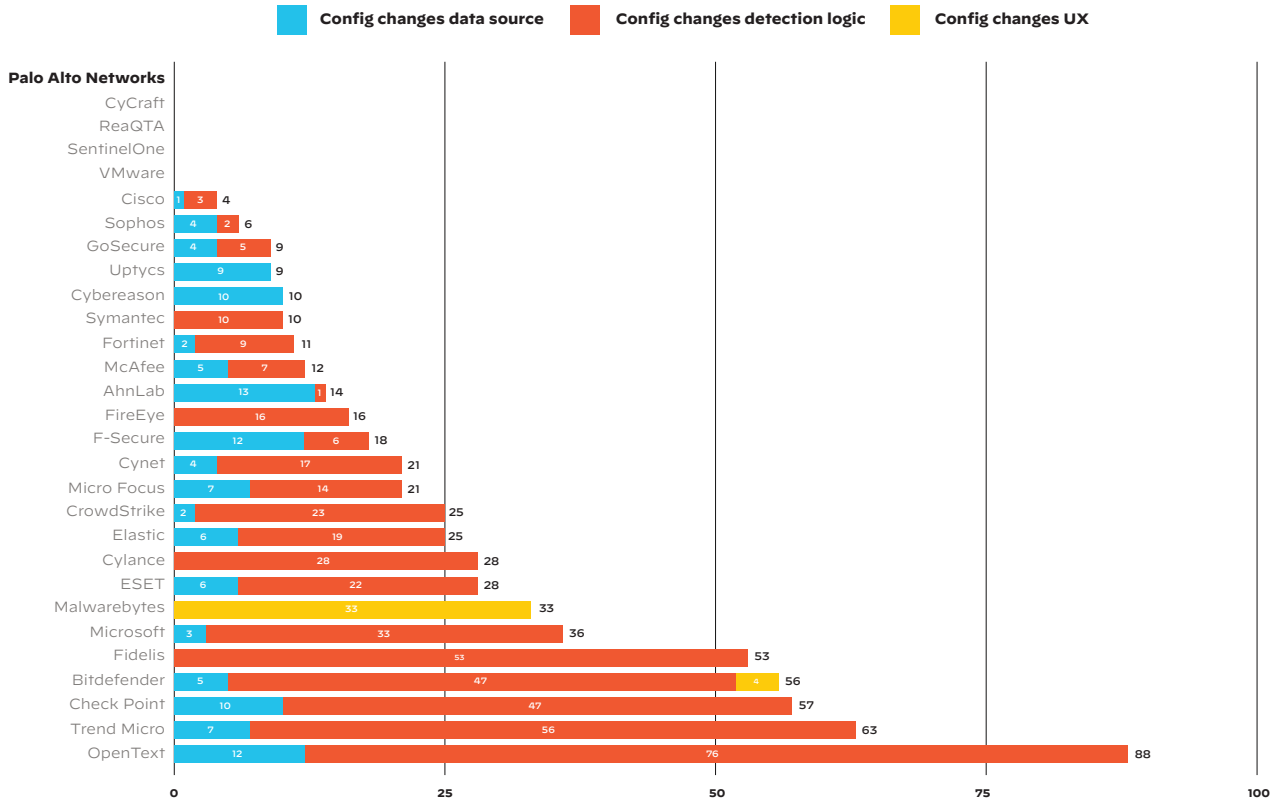

Figure 5: Number of configuration changes per vendor in the Round 3 evaluation

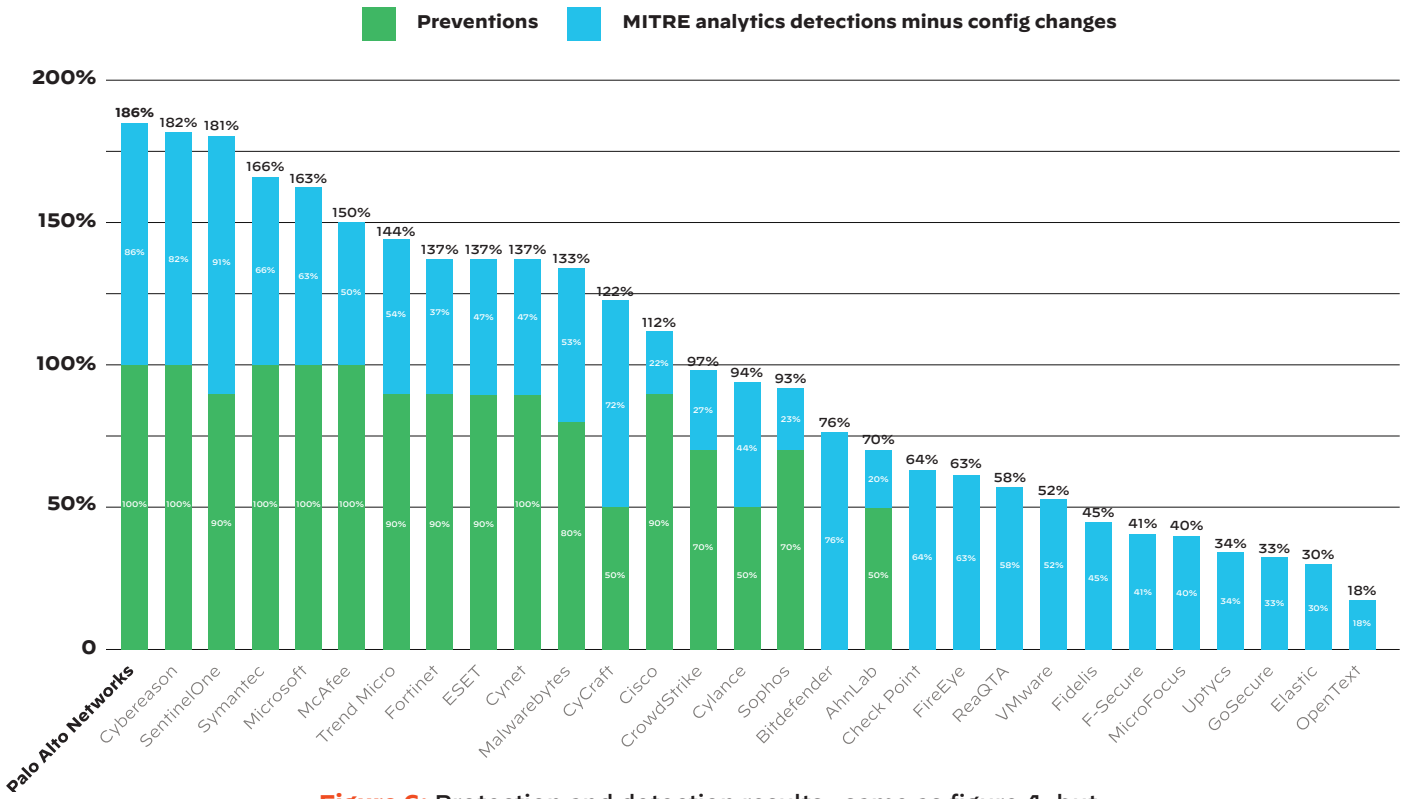

Figure 6: Protection and detection results—same as figure 4, but detection results achieved with a Configuration Change are not included

The Numbers Don't Tell the Whole Story

When examining the MITRE results, it's important to look at the product screenshots to get a better sense of the story that is being told to the security analyst.

For example, Cortex XDR is the only unified solution that has the ability to correlate and stitch together network data, endpoint data, and third-party data in one causality view, and then apply analytics to identify anomalies on that stitched data story. We can see an example of this in the Cortex XDR screenshot from step (figure 7) 5.C.2 in the evaluation. In this step, we observe a detailed kill chain with a cross-host view that shows when the attacker performed lateral movement from Linux to Windows via Server Message Block (SMB). The view shows data from both endpoint and network sources stitched together. Cortex XDR detected the technique leveraging our analytics engine on the combined data story.

Another example is shown in step 17.A.6 of the evaluation (figure 8), where we see code injection via Remote Procedure Call (RPC) within an encrypted HTTPS session. This result is achieved despite the encrypted channel by monitoring RPC as a data source, despite masquerading via a port change. This results in visibility into the command-and-control (C2) connection. This level of transparency was enabled by the detailed monitoring of RPC via the Cortex XDR agent.

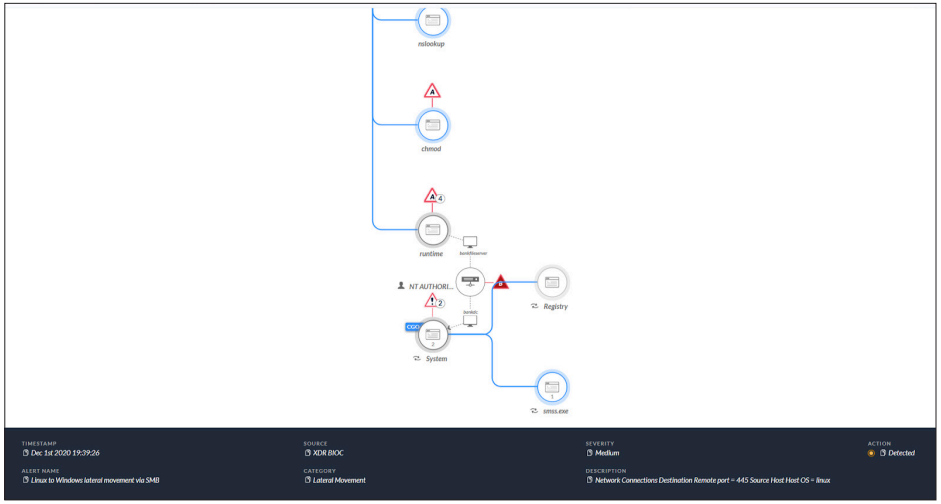

Figure 7: In step 5.C.2, a detailed kill chain with a cross-host view

CORTEX XDR BY PALO ALTO NETWORKS

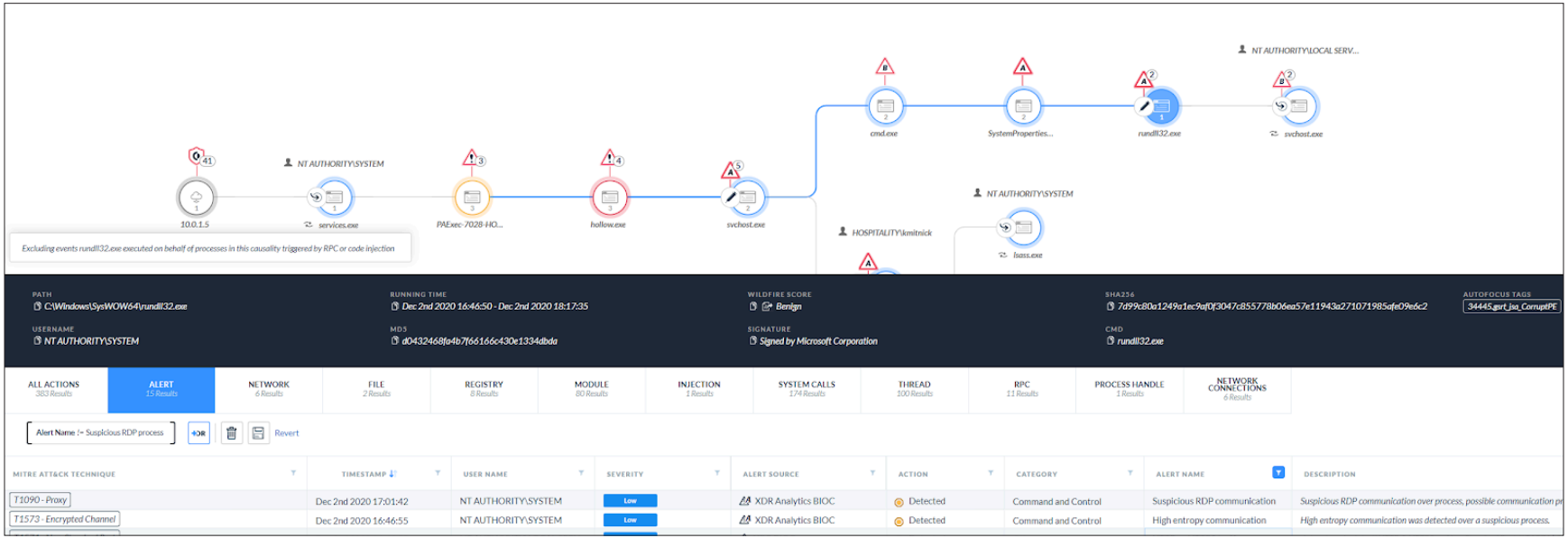

Figure 8: In step 17.A.6, Cortex XDR tracking RPC calls within HTTPS despite masquerading

CORTEX XDR BY PALO ALTO NETWORKS

Because Cortex XDR integrated endpoint data with network data that contained App-ID™ information, it was able to identify when the MITRE Red Team performed lateral movement using SSH between the Windows and Linux hosts. This is seen in step 5.B.1 of the evaluation (figure 9), where Cortex XDR offered visibility into how the attacker performed lateral movement, revealing which protocols were used.

Through examples like these, we can see that the depth of value provided by Cortex XDR goes far beyond the detection numbers provided in the evaluation results. Cortex XDR provides a thorough and complete representation of the attack intricacies by pulling in data from multiple data sources and leveraging our analytics engine to stitch together a complete causality view for the administrator without the need for manual investigation and correlation.

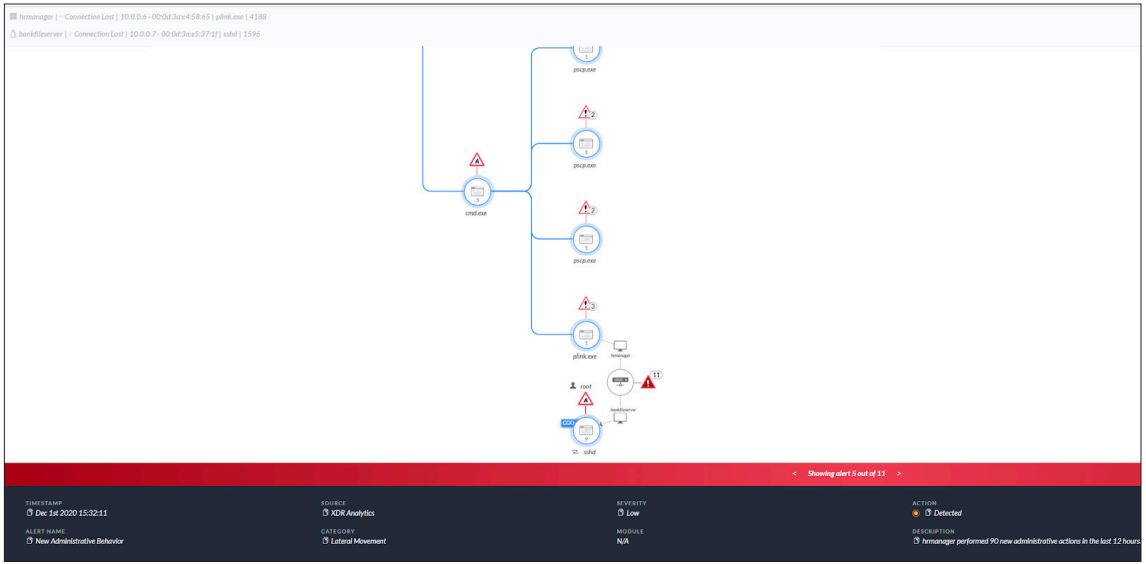

Figure 9: In step 5.B.1, Windows to Linux lateral movement over SSH

CORTEX XDR BY PALO ALTO NETWORKS

Can't Get Enough of Round 3? We Have More!

For security teams trying to make sense out of the MITRE ATT&CK results, we understand the challenge of deciphering the various vendor interpretations. In addition to the details provided here, we encourage you to read this blog post by Palo Alto Networks Field CTO Josh Zelonis: Don't Let Vendor Exuberance Distract from the Value of the MITRE ATT&CK Evaluation. It should help you understand some of the more nuanced scrutiny, including further explanation of the metrics of visibility and analytics.

If you are interested in learning more about the attack scenarios emulated in this evaluation and the technologies that best protect and detect these techniques, register for our on-demand webinar Carbanak+FIN7: MITRE ATT&CK Results Unpacked.

EDRs Are on a Fast Track to XDR, So Hold on to Your Hats

Curious to learn more about Extended Detection and Response (XDR) as it gains traction in the marketplace? Download our e-Book *XDR: Extended Detection and Response* to learn more including:

- Challenges with the current state of detection and response
- Tactical use cases for improving security operations with XDR
- The definition and key requirements of XDR

More About MITRE

For further information on the ATT&CK Framework, visit MITRE.org. Check out the ATT&CK Navigator tool to help you navigate, annotate, and visualize ATT&CK techniques.

About MITRE Engenuity

MITRE Engenuity ATT&CK Evaluations are paid for by vendors and are intended to help vendors and end-users better understand a product's capabilities in relation to MITRE's publicly accessible ATT&CK® framework. MITRE developed and maintains the ATT&CK knowledge base, which is based on real world reporting of adversary tactics and techniques. ATT&CK is freely available and is widely used by defenders in industry and government to find gaps in visibility, defensive tools, and processes as they evaluate and select options to improve their network defense. MITRE Engenuity makes the methodology and resulting data publicly available so other organizations may benefit and conduct their own analysis and interpretation. The evaluations do not provide rankings or endorsements.

MITRE ENGENUITY™

A Foundation for Public Good

Reference

1. "Detection and Protection Categories," ATT&CK Evaluations, MITRE Engenuity, last visited May 21, 2021, https://attackevals.mitre-engenuity.org/enterprise/carbanak_fin7/#detection-categories.
2. Ibid.
3. Ibid.

paloalto® NETWORKS

3000 Tannery Way
Santa Clara, CA 95054

Main: +1.408.753.4000
Sales: +1.866.320.4788
Support: +1.866.898.9087

www.paloaltonetworks.com

cortex_eb_essential-guide-mitre-round-3-052621