The Essential Guide to the 2023 MITRE Engenuity ATT&CK Evaluations

Enterprise Edition

This e-book provides a comparative look at how endpoint security solutions performed in the most recent ATT&CK evaluation, with guidance on how to explore the results further. We include key descriptions of MITRE's testing methodology, the tools MITRE Engenuity provides to help visualize and compare results, and considerations for analysis to help you assess which vendor best fits your organization's endpoint security needs.

paloalto® NETWORKS | CORTEX® BY PALO ALTO NETWORKS

Table of Contents

Introduction

Since 2018, the ATT&CK® Evaluations have provided the industry's most sophisticated public attack simulations for security vendors to essentially "test their wares" against attack methodologies representative of real-world threats.

Focused on the technical ability to address known adversary tactics, techniques, and procedures (TTPs), the evaluations provide the opportunity to analyze endpoint detection and response (EDR) products against real-world attack sequences.

This year's evaluation was broken down into two detection-only scenarios (figure 1), named Carbon and Snake, referring to tools created and used by the Turla threat group. A protection test followed, mirroring the techniques in the detection tests, with enough entropy injected not to look identical to the detection test. The two detection scenarios each had a combined total of 19 steps consisting of multiple substeps that map to actual techniques in the MITRE ATT&CK framework. Combined, there were *143 substeps* that each vendor had the opportunity to detect. For each of these substeps, the MITRE Engenuity team recorded whether each solution had a detection for the action taken.

In this year's evaluation, ***only*** Palo Alto Networks prevented every step in the protection scenario and delivered an analytic detection in every substep of the two detection scenarios.

At a high level, Cortex XDR achieved the following against the TTPs used by Turla:

- 100% block rate in the protection scenario
- 100% analytic coverage in the detection scenarios
- 99.3% technique-level detections (142/143)
- 100% visibility

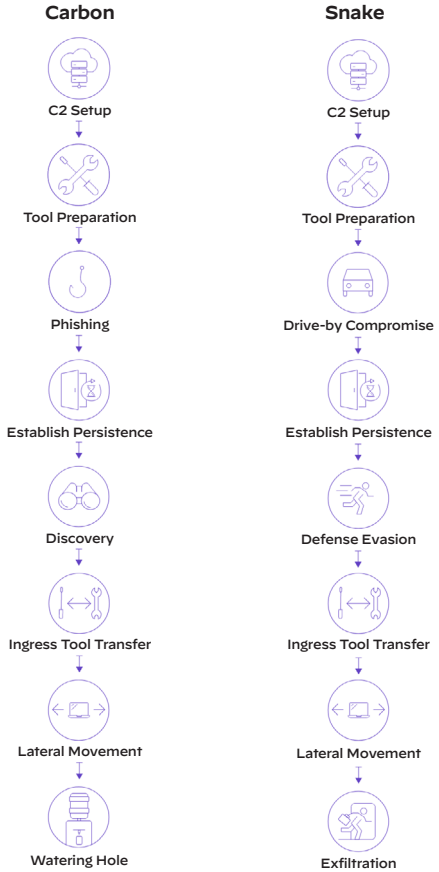


Figure 1: Turla operational flow

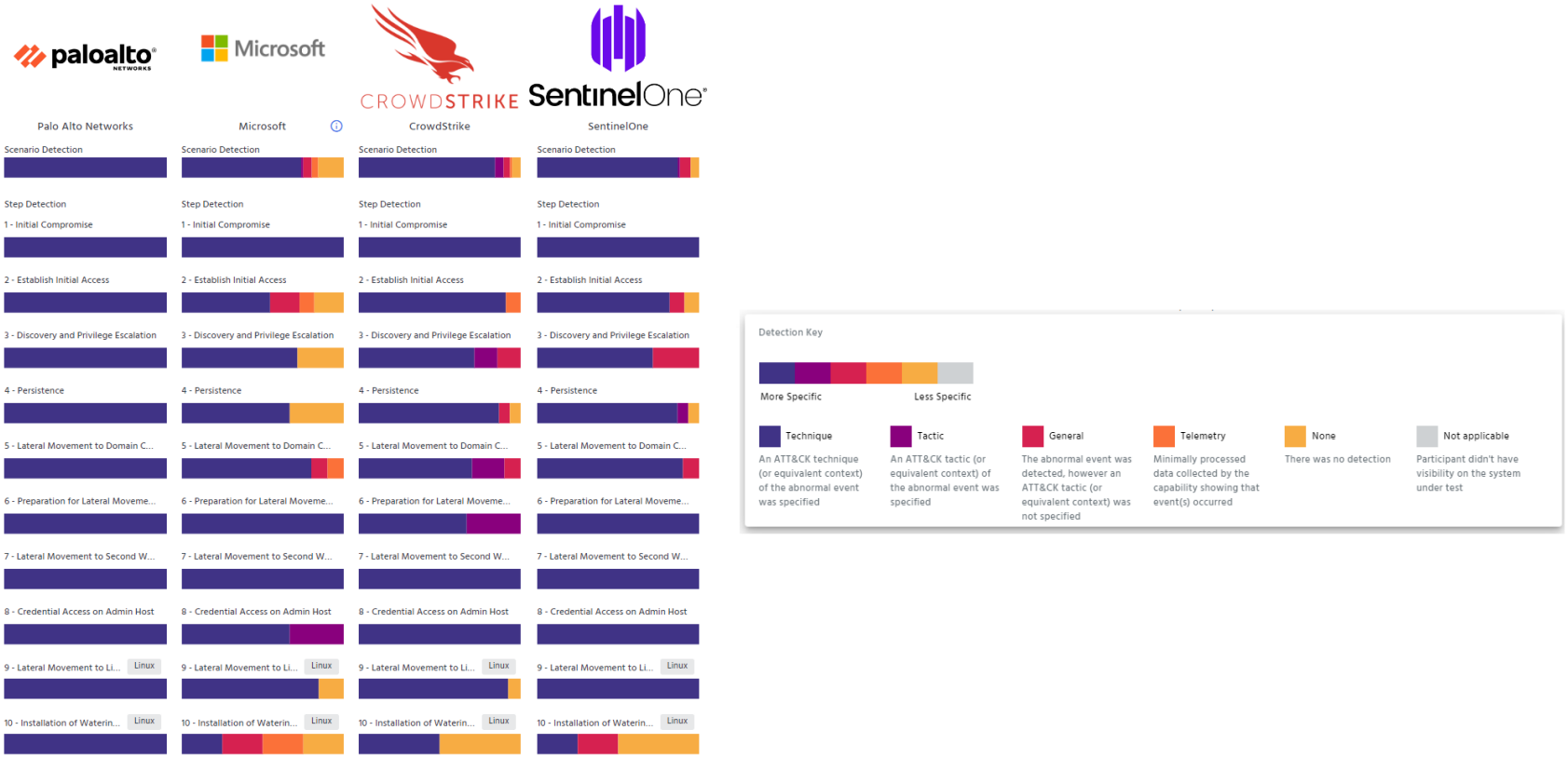


Figure 2: Results for the first detection scenario as seen on the MITRE Engenuity ATT&CK Evaluation website shows the 2023 Turla evaluation results for the first detection test

Evaluations Overview

In the last year's evaluation, MITRE Engenuity emulated the TTPs of two russia-based threat groups, Wizard Spider and Sandworm, known for their high volume of financially motivated and destructive attacks. This year, MITRE Engenuity chose to emulate Turla, an organization within Russia's FSB, know for its stealthy attacks and targeted intrusions. For Round 5, the evaluation continued to draw a large number of vendors (29 this year) which highlights the importance of this evaluation for the industry.

As described by MITRE, Turla (aka Pensive Ursa) targeted victims in over 45 countries in a wide range of sectors, including government entities, embassies, and military organizations, education, research, and pharmaceutical companies. In addition, this threat group had an active part in the Russian-Ukraine conflict that started in February 2022. According to the Ukraine CERT, Turla leveraged espionage attacks against Ukrainian targets, specifically against the Ukrainian defense sector. While Turla mainly used their espionage arsenal to target Windows machines, the group also has tools that can attack macOS and Linux machines.

> ***This round followed Turla through a multi-phased, intelligence-collection campaign. The emulation highlights how Turla achieves post-exploitation persistence with a minimal footprint through in-memory or kernel implants, evades detection by defensive tools, and exfiltrates sensitive information from Linux and Windows infrastructure.***
>
> –MITRE

Turla is an extremely well funded and advanced threat group, and while their techniques are used to target government agencies, the advancements they have made are often copied by numerous financially motivated threat groups targeting a wider number of industry verticals.

These evaluations assess 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.

What was the motivation behind the ATT&CK evaluations?

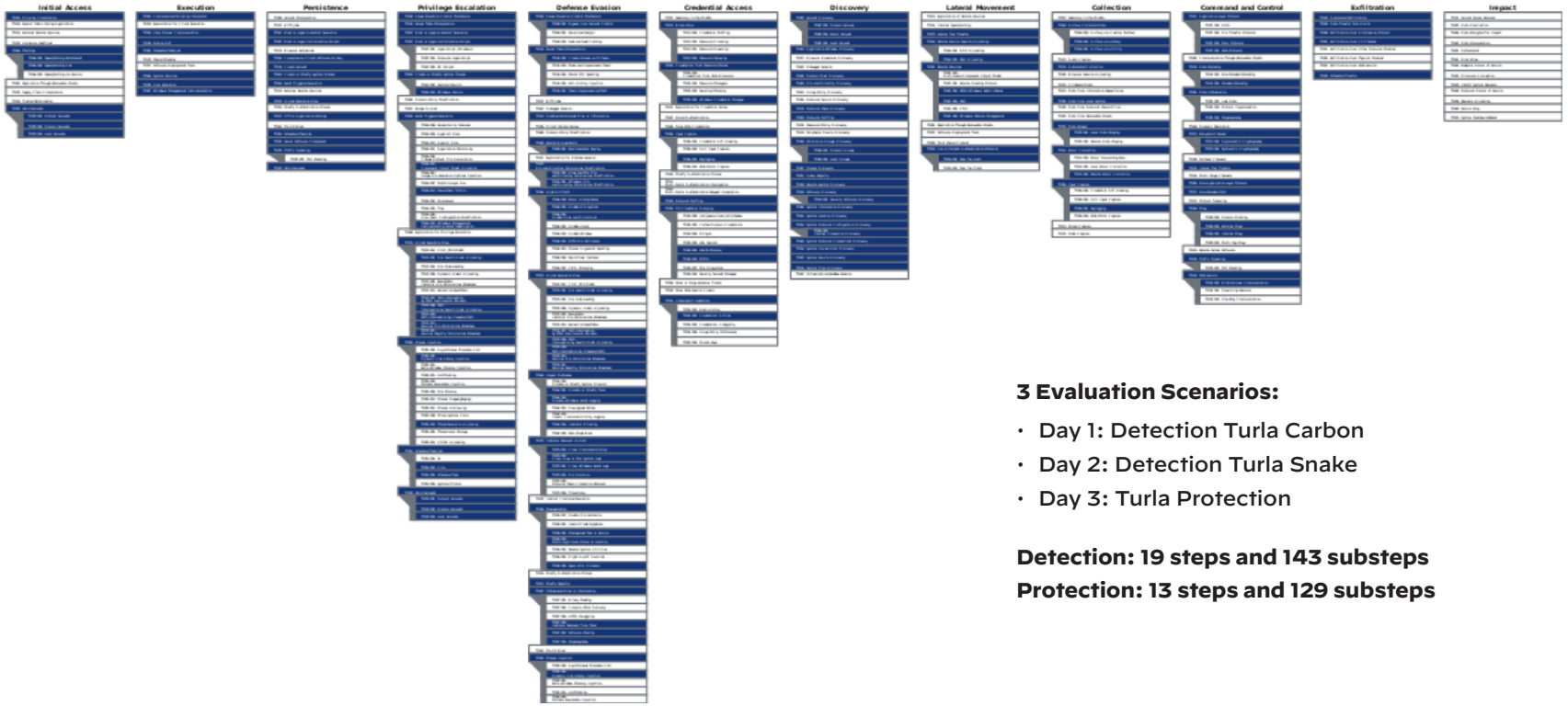
- Vendors are using ATT&CK to articulate their capabilities, but there is no neutral authority to evaluate their claims.

What are the ATT&CK evaluations?

- Open, transparent and objective. Methodology and results are published openly and clearly.
- Evaluates both protection and detection efficacy. (Protection evaluation was included starting in Round 3.)
- A compilation of the detections MITRE Engenuity observes in response to an emulated adversary's tactics and techniques.

What aren't the ATT&CK evaluations?

- Not designed to address noise or false positives.
- Not a ranking or rating of a vendor's technology.



3 Evaluation Scenarios:

- Day 1: Detection Turla Carbon
- Day 2: Detection Turla Snake
- Day 3: Turla Protection

Detection: 19 steps and 143 substeps
Protection: 13 steps and 129 substeps

Note: The items in blue are the techniques in the MITRE ATT&CK Enterprise framework that were emulated.

Figure 3: The MITRE ATT&CK framework: Turla. Explore in ATT&CK Navigator.

The MITRE ATT&CK Framework

- The MITRE ATT&CK framework has become the standard for how the security world communicates about adversaries and their techniques.
- **ATT&CK** stands for Adversarial Tactics, Techniques, and Common Knowledge.
- Provides detailed information about all the adversarial techniques.
- Details of threat groups that have used these techniques.
- Useful information about how to detect and mitigate these tactics and techniques.

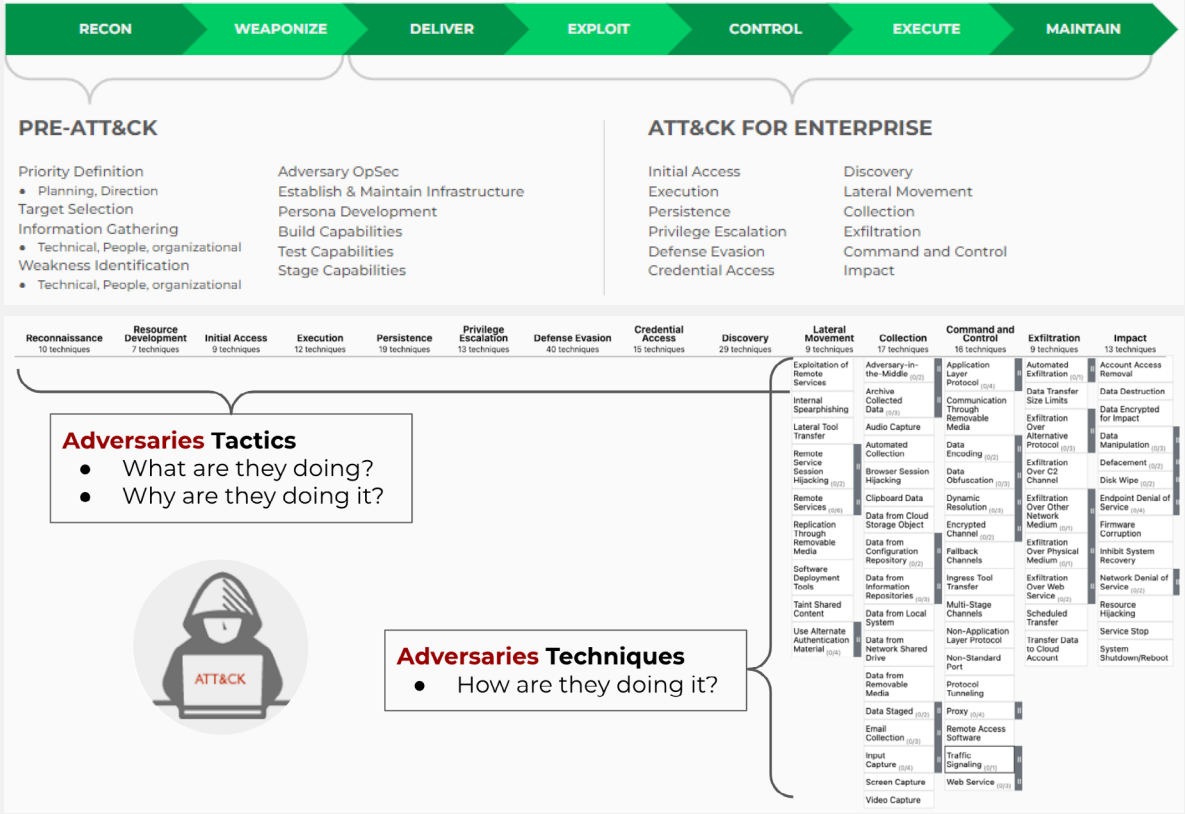


Figure 4: Understanding the MITRE ATT&CK framework

MITRE Engenuity's Approach

Focused on articulating how detections occur rather than assigning scores to vendor capabilities, MITRE Engenuity categorizes each detection based on quality and precision. (See Detection Categories for more details.) While MITRE Engenuity makes every effort to capture different detections, vendor capabilities may be able to detect procedures in ways that MITRE Engenuity 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 substep does not mean it applies to all techniques of that step. For proof of detection in each category, MITRE Engenuity 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 Engenuity reviews the screenshots provided, the notes taken during the evaluation, the results of follow-up questions to the vendor, and vendor feedback on draft results.

> *"To provide transparency around the ability of defensive solutions to address the behaviors described in ATT&CK and propel the enterprise security market forward, the Enterprise Evaluations methodology was specifically designed to be data-driven and focus on this very specific topic."*
>
> –Frank Duff, Ex-Director of ATT&CK Evaluations

They also independently test procedures in a separate lab environment and review open-source tool detections and forensic artifacts. This testing informs what is considered to be a detection for each technique.

Using MITRE Engenuity to Help Evaluate Endpoint Security Solutions

For organizations reviewing EDR solutions and vendors, the MITRE Engenuity results compare the various levels of security efficacy by participating vendors, all aligned around a common lexicon to ensure parity and continuity across the evaluations.

So, how can the evaluations help inform a defensive strategy for solution providers like us? At Palo Alto Networks, participating in these evaluations allows us to be tested by a neutral, unbiased third party, leveraging current, real-life sophisticated attack sequences. This method of testing yields constructive insights into how we can build more effective detection and prevention solutions.

In using the modern attack TTPs from groups such as Turla and emulating the attack scenarios 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 solution or product modifications and guide fine-tuning any steps that may have underperformed.

MITRE Engenuity Round 5 Methodology

Environment

The evaluations were performed using Microsoft Azure cloud services.

NOTE: The use of cloud services is solely for efficiently provisioning and managing resources. Evaluations environments should be considered as if they were an "On-Premises" environment.

There were two scenarios with separate networks and domains, with Windows Defender disabled for certain portions of the evaluations. The evaluation networks contained domain joined instances running Windows Server 2019, Windows 10 Professional, as well as instances running Ubuntu 20.04 LTS. The specific versions in scope for the evaluation are listed below:

- Windows Server 2019
  - » Publisher: MicrosoftWindowsServer
  - » Versions: 2019.0.20190410, 17763.3406.220909
  - » SKU: 2019-Datacenter
- Windows 10 Professional
  - » Publisher: MicrosoftWindowsDesktop
  - » Versions: 18362.1256.2012032308, 19044.2006.220909
  - » SKUs: 19h1-pro-gensecond, win10-21h2-pro-g2
- Ubuntu 20.04 LTS
  - » Publisher: Canonical
  - » Version: 20.04.202207130
  - » SKU: 20_04-lts-gen2

Vendors were provided VPN profiles to connect to the evaluation networks.

Learn more about the environment here.

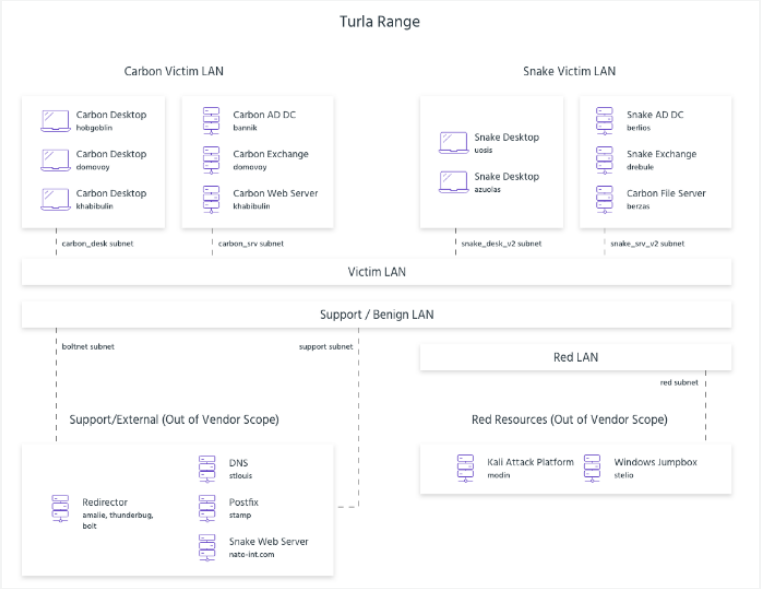


Figure 5: Turla evaluation environment

Detection Categories

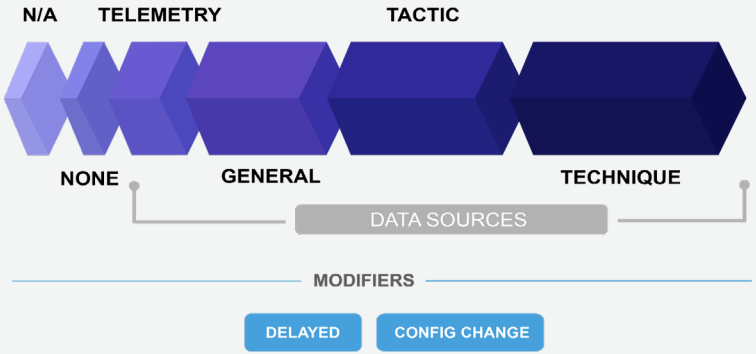


Figure 6: Turla detection categories

Vendors use their own terminology and approaches to detect and protect potential adversary behavior. They provide this information to us in their unique way, and then it is our responsibility to abstract the data using categories to talk about the products in similar ways.

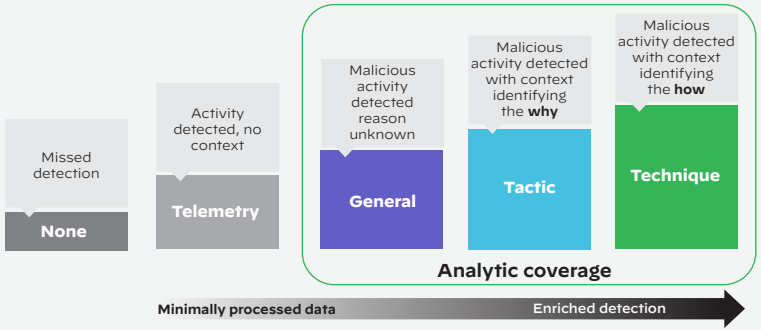


Figure 7: MITRE Engenuity detection categories

These categories are divided into two types: “Main” and “Modifier.” Each detection or protection receives one main category designation, which relates to the amount of context provided to the user, and may optionally receive one or more modifier category designations that help describe the event in more detail. For the Turla evaluation, there are six main detection categories representing the amount of context provided to the analyst, and three main protection categories.

Learn more about detection categories here.

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.
- **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.

Detection-Type Modifiers

MITRE Engenuity 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. This approach allows end users of the results to determine what they value most in a detection (e.g., some organizations may want telemetry, while others would want Technique detection).

- **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.

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.

Cortex XDR vs. Turla

This year marks the fifth annual evaluation, and the MITRE Engenuity red team focused on emulating the methods of Turla, a threat group our Unit 42 threat researchers have studied extensively. Turla is an extraordinarily well-funded and sophisticated Russia-based threat group that has infected victims in over 45 countries. They have targeted government agencies, military groups, diplomatic missions, as well as research and media organizations. Turla's infamy stems from its covert exfiltration tactics, including water holing of government websites, custom rootkits, elaborate command-and-control network infrastructure, and deception tactics. In speaking with defenders who participated, it is clear that MITRE Engenuity took a great leap forward in the sophistication of their attack methods this year.

The blue team deployed Cortex XDR Pro for Endpoint agent on both Windows and Linux endpoints. No additional solutions were deployed, and Cortex XDR was configured with default settings as it would be out of the box, with the only changes enabling the quarantining of malicious files and, for Linux, enabling the option to treat grayware as malware.

The Cortex XDR Difference—the Data Doesn't Lie

Though just about every major endpoint security vendor is claiming to deliver 100% protection and detection, the data provided by MITRE Engenuity tells a different reality. In this year's evaluation, **only Cortex XDR provided 100% Protection** while delivering 100% **Visibility** and 100% **Analytic Coverage** (detections) with zero configuration changes or delayed detections.

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 an 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.

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

The purpose of these evaluations is to provide insight into three capabilities:

1. **Visibility**: What can a solution see?
2. **Detection**: What actions can a solution accurately identify as malicious?
3. **Protection**: What malicious actions can a solution prevent?

In addition, the quality of our detections is unparalleled, with 142 of 143 detections as technique-level detections—*the highest-quality detection possible*. The one other detection was recognized as a tactic-level detection. Every one of the 129 substeps in the Protection evaluation was blocked. All of this was accomplished with *zero configuration changes* and *zero delayed detections*.

In fact, if we exclude detections resulting from a configuration change, Cortex XDR was the only vendor with no missed detections (detection type None). In other words, Cortex XDR was the *only one with 100% visibility*.

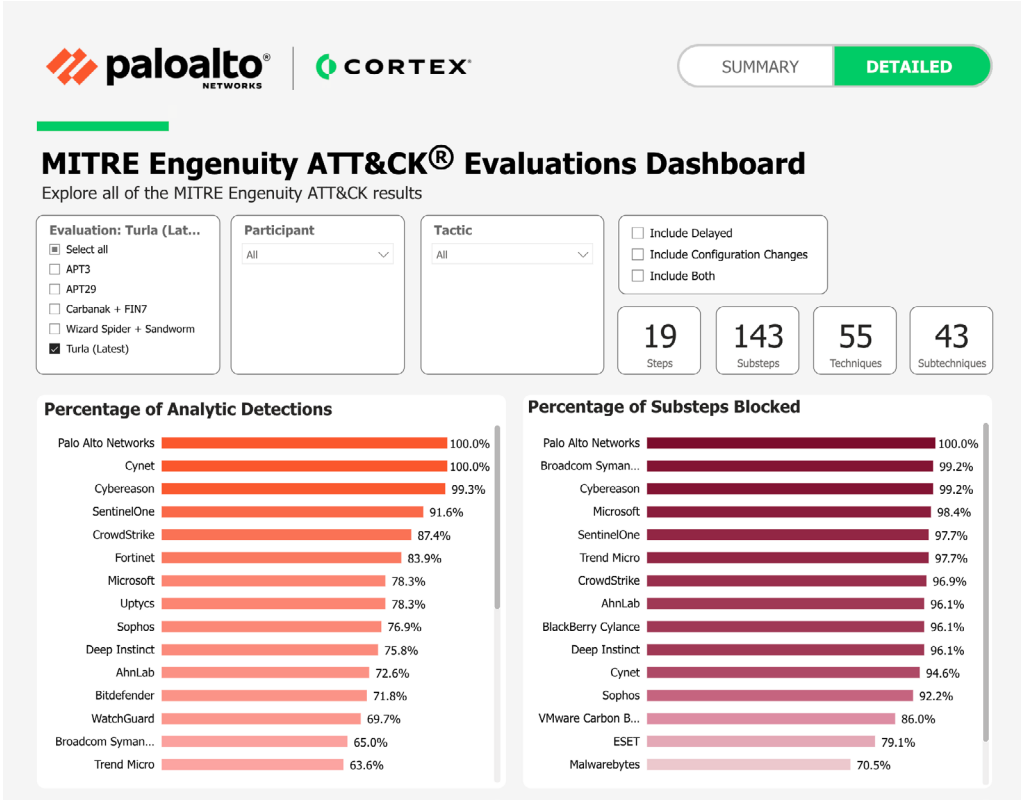


Figure 8: The Palo Alto Networks MITRE Engenuity ATT&CK Evaluations Dashboard. This snapshot shows that Cortex XDR provided the highest number of Technique detections and was the only solution to block 100% of attack substeps in the Protection test.

The results from this year's evaluation serve to reflect the tremendous amount of effort that Palo Alto Networks continues to pour into both adversarial research and engineering in endpoint security, putting that knowledge to work to help our customers remain safe in the face of an increasingly hostile cyber world.

About Configuration Changes

MITRE Engenuity allows for solution providers to have a *do-over day* to achieve a better detection after the initial test was executed. Detections observed during the do-over day are noted by the configuration change modifier. 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 to garner a better result. MITRE Engenuity provides this opportunity for vendors to have the chance to validate how changes to the solution may improve security efficacy.

In the real world, when an attack is not prevented, the attacker does not give you a second chance with an opportunity to introduce a change in security configuration. While we understand and appreciate the intent for configuration changes in these evaluations, we feel it is more realistic to exclude detections directly resulting from a configuration change when comparing results. Unfortunately, many vendors are touting industry-leading results while including detections achieved in the do-over resulting from a configuration change. It is important to note that there is no limit to what can be changed when making a configuration change, and there is also no commitment from the vendor to include these changes in their production code.

Examples of configuration changes include:

- A new rule is created, a preexisting 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.

Number of Delayed Detections

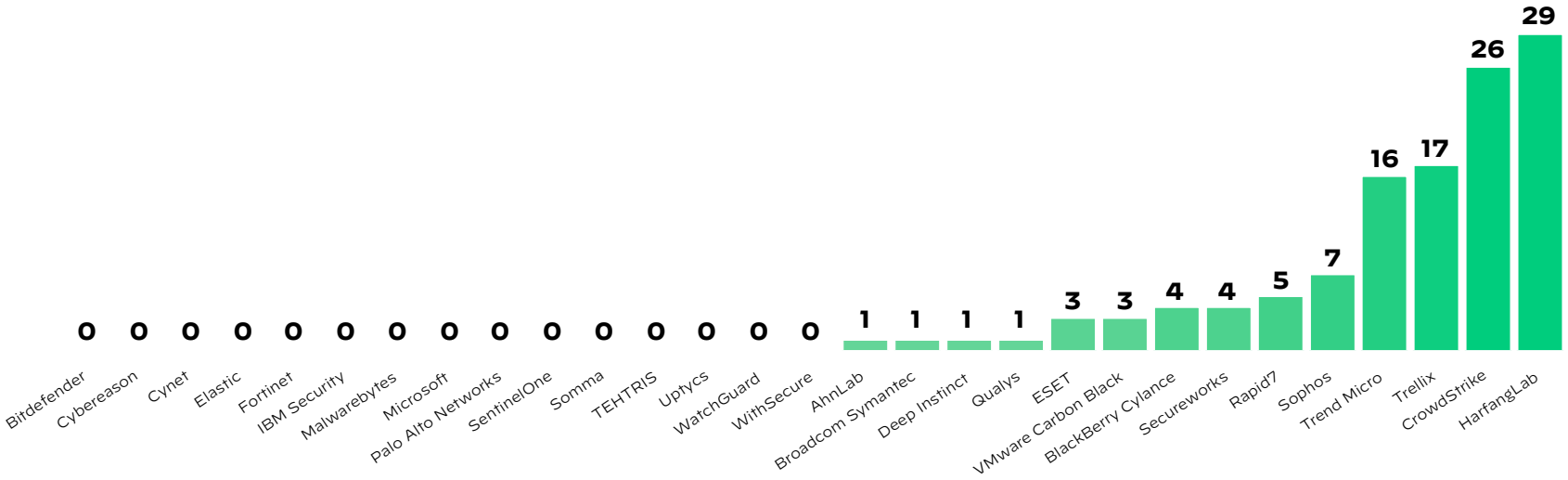


Figure 9: The charts above show the number of delayed detections and the number of detections, which resulted from configuration changes that were observed in the do-over day

Detections Resulting from a Configuration Change

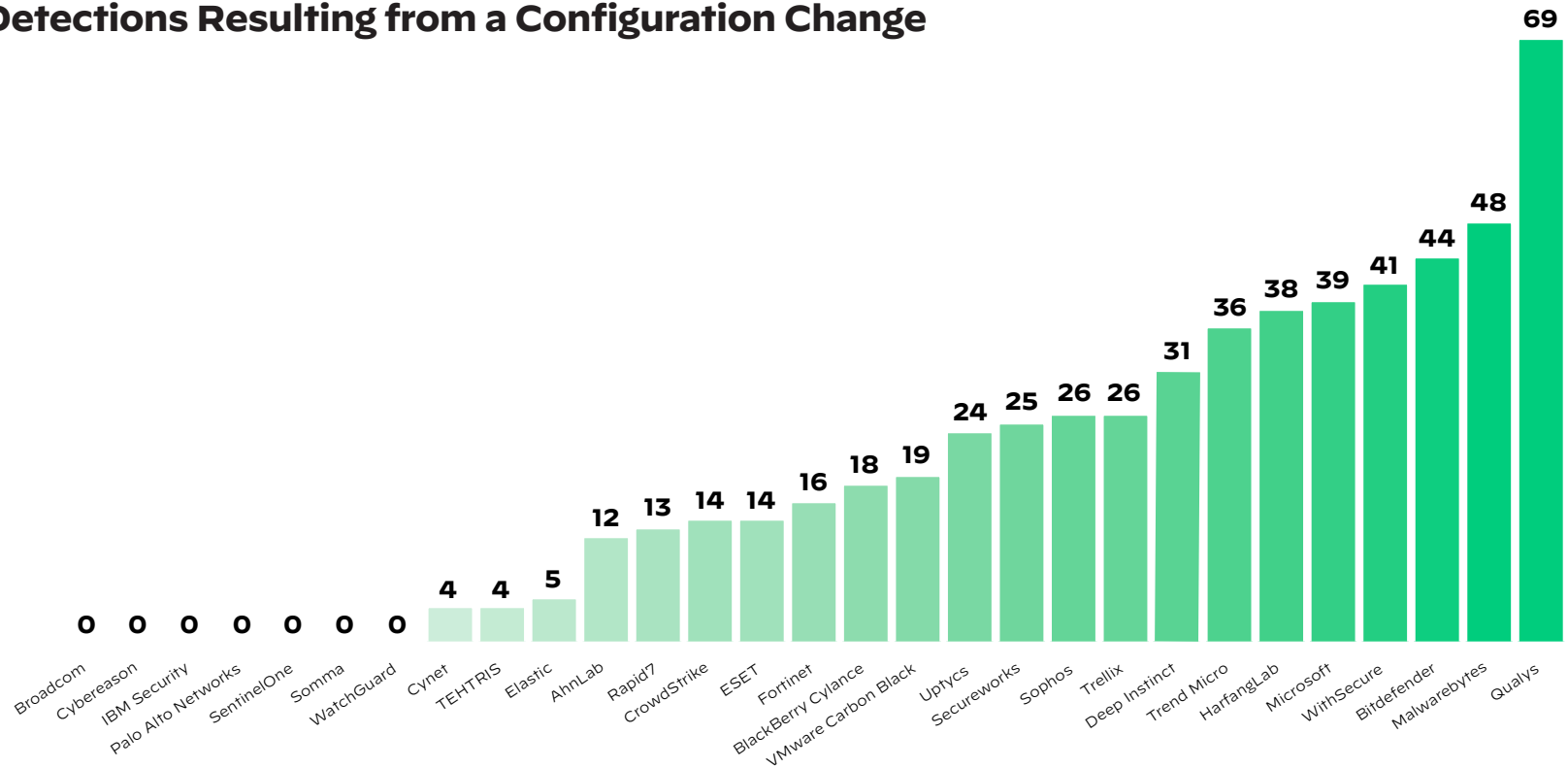


Figure 9: The charts above show the number of delayed detections and the number of detections, which resulted from configuration changes that were observed in the do-over day (continued)

Configuration Changes Excluded

When all detections resulting directly from configuration changes are excluded from the results tallies, Palo Alto Networks leads no matter how we view the data.

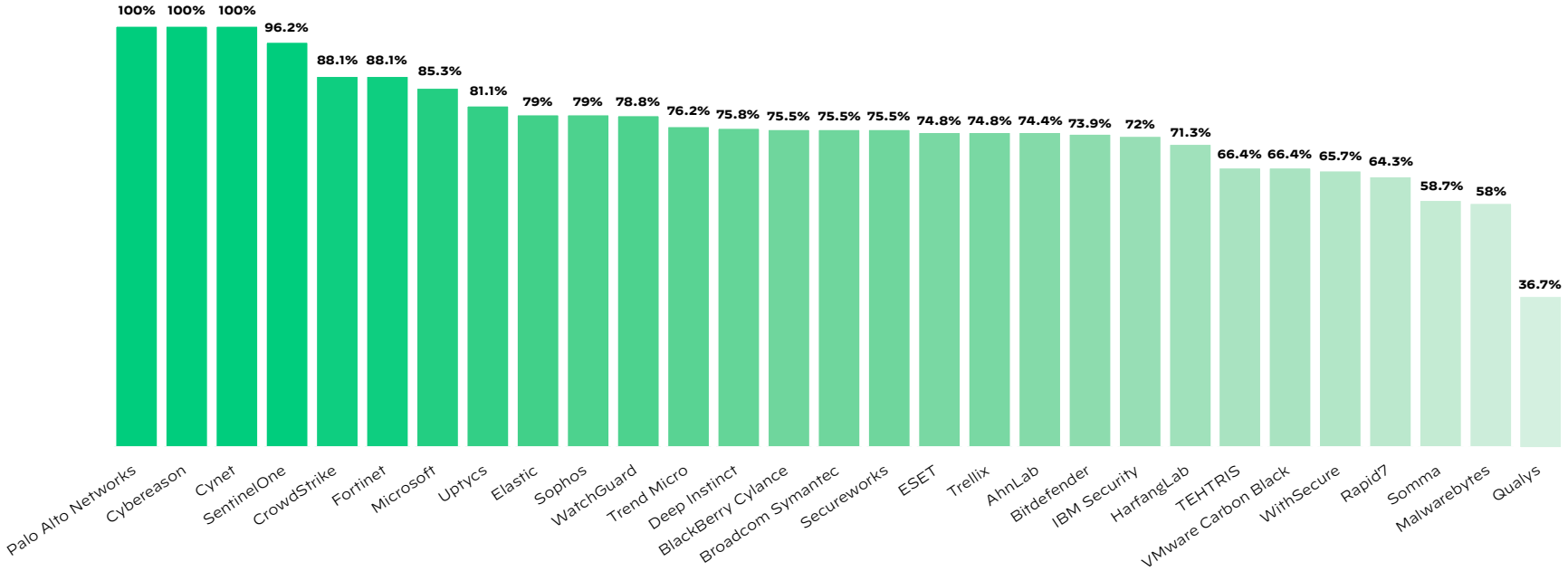


Figure 10: Visibility is the foundation for preventions and detections. Cortex XDR was unbeaten in attack visibility, noted here as "unique detections per substep."

Percentage of Analytic Detections

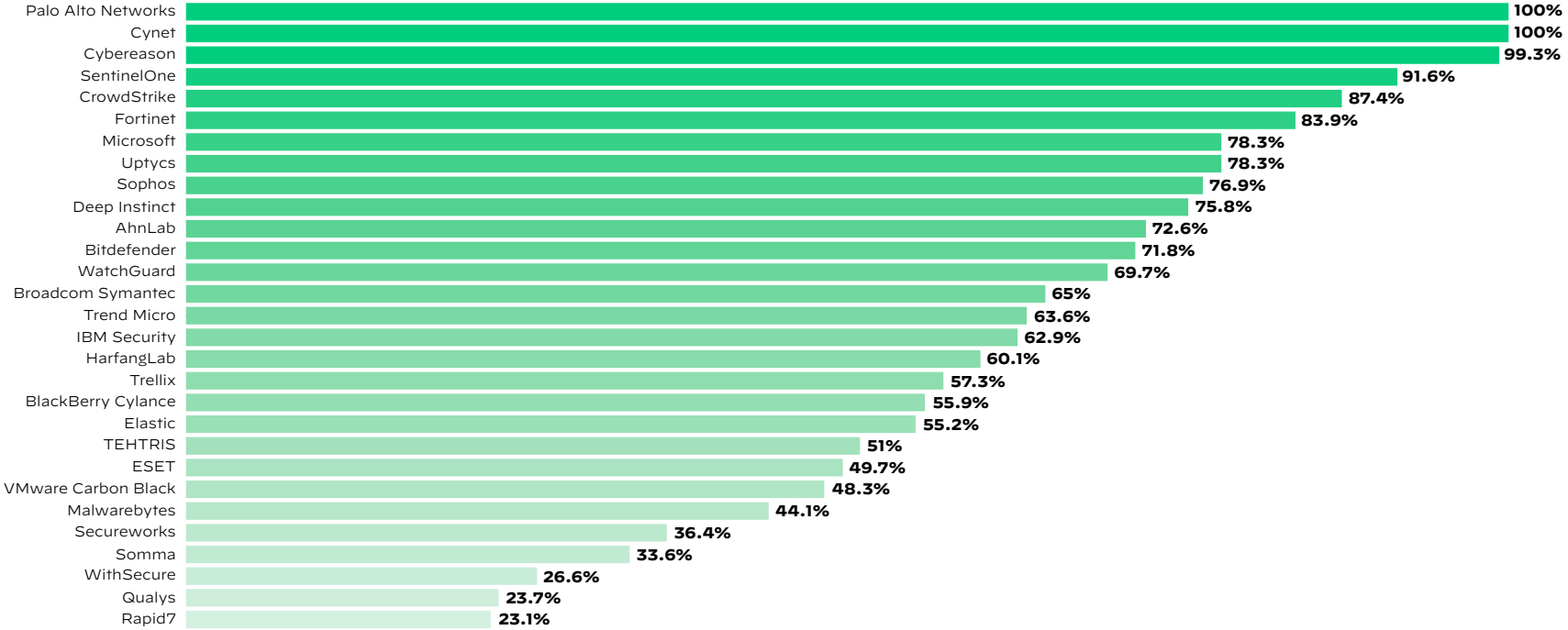


Figure 11: Cortex XDR delivered high-quality detections for every malicious action in the detection phase, notching a 100% Analytic Detection rate

Percentage of Technique-Level Detections

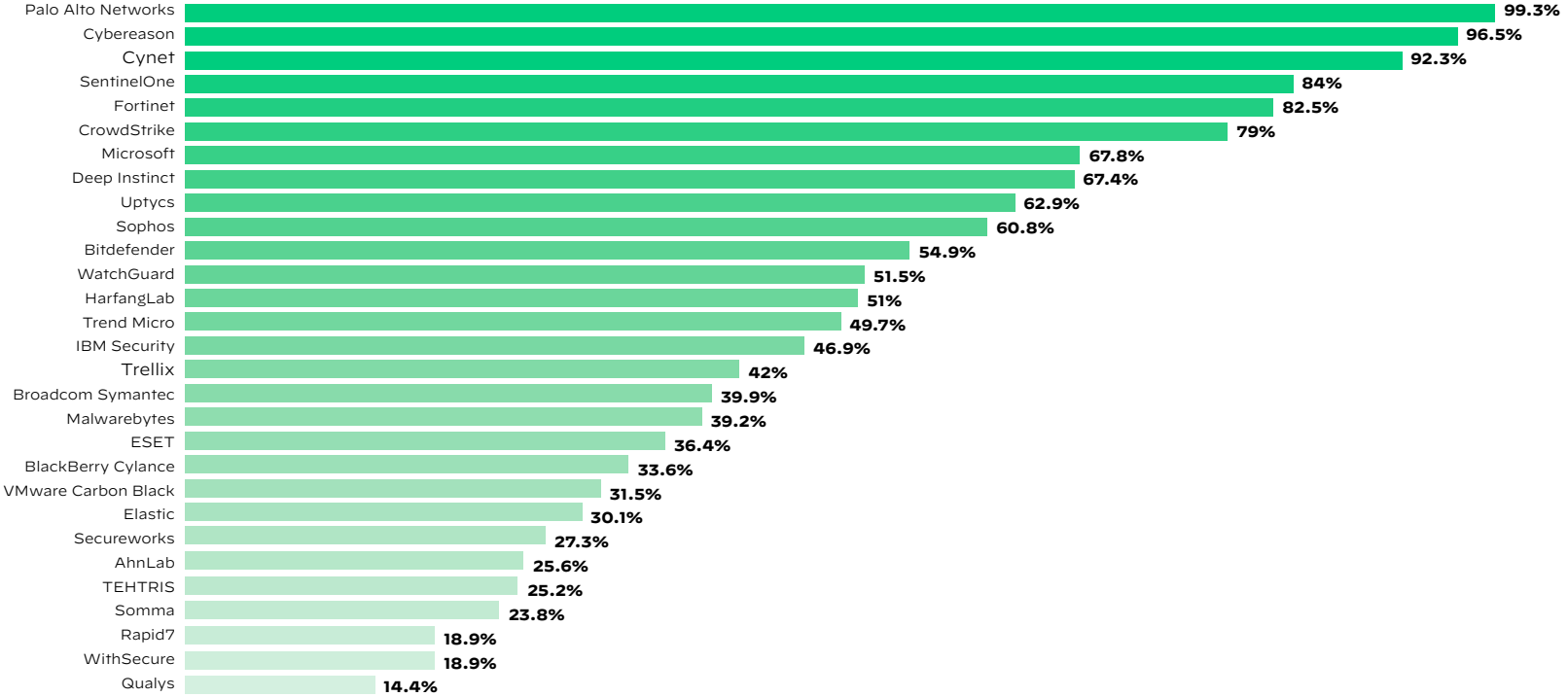


Figure 12: In addition to providing 100% Analytic Detections, Cortex XDR provided the highest number of quality detections with 142 of 143 being technique-level detections

Percentage of Technique-Level Detections by Leading EDR Vendors

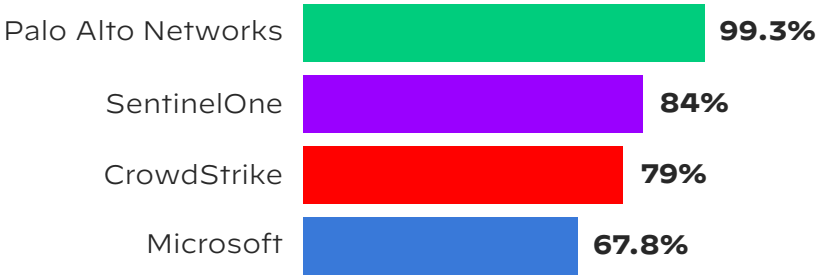


Figure 13: Percentage of technique-level detections by leading EDR vendors

Percentage of Substeps Blocked by Leading EDR Vendors

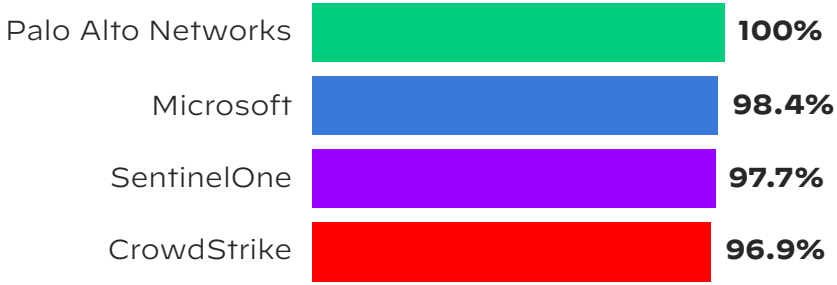


Figure 14: Percentage of substeps blocked by leading EDR vendors in the Protection test

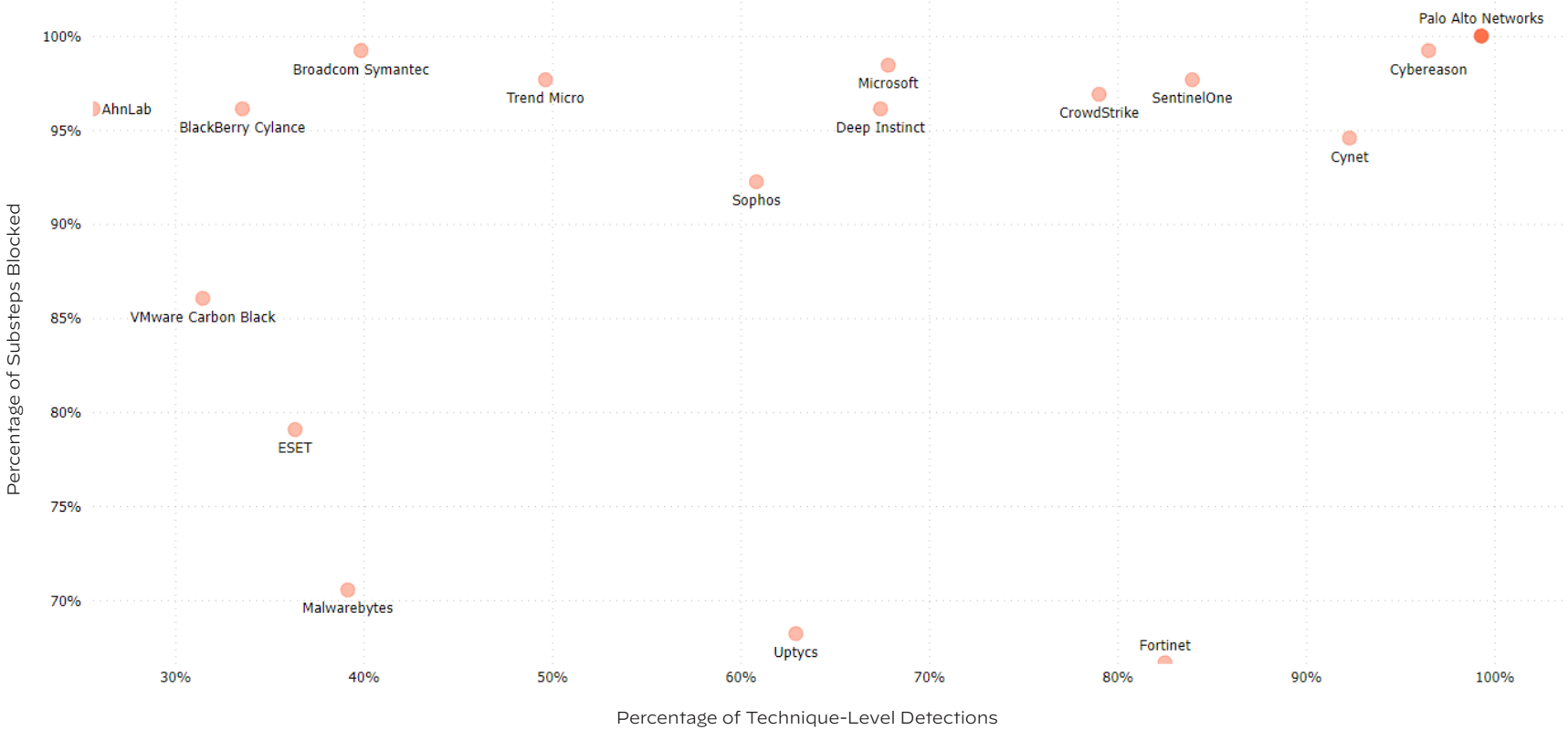


Figure 15: Combined protections and detections graph with all protection substeps included, and detections resulting from configuration changes excluded

Conclusion

As we reflect on these evaluations, it's clear they provide valuable insights into the capabilities of security solutions. They allow organizations to make informed decisions about their endpoint security needs, guided by an independent third-party evaluation highlighting each solution's ability to detect and/or prevent a wide range of attack techniques. At Palo Alto Networks, participating in these evaluations underscores our commitment to delivering the best possible detection and prevention solutions in the face of evolving cyberthreats.

We invite you to explore the detailed results and insights provided in our MITRE Engenuity ATT&CK Evaluations Dashboard. As we continue to adapt to the ever-changing threat landscape, Palo Alto Networks remains dedicated to helping our customers stay safe in an increasingly hostile cyber world.

More About MITRE ATT&CK and Cortex XDR

If you're interested in learning more about the attack scenarios emulated in this evaluation and how Cortex XDR performed, we have a variety of resources available on demand:

2022 Results

- All about our results in under three minutes. Watch the video.
- View our on-demand webinar. Dissecting the 2022 MITRE Engenuity ATT&CK Evaluations.
- Visit our webpage and read our 2022 MITRE Engenuity ATT&CK Evaluations Results blog for more information.

2023 Results

- Read the 2023 MITRE Engenuity ATT&CK Evaluations Results blog.
- Explore our new MITRE Engenuity ATT&CK Evaluations Dashboard. See our stellar results from the past five years.
- Watch Cortex XDR take down cyber adversary Turla in this retro, arcade-style action video inspired by the 2023 MITRE Engenuity ATT&CK Evaluation.

About the MITRE Engenuity ATT&CK Evaluations

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.

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.

MITRE ENGENUITY™

A Foundation for Public Good

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

© 2023 Palo Alto Networks, Inc. Palo Alto Networks and the Palo Alto Networks logo are registered trademarks of Palo Alto Networks, Inc. A list of our trademarks can be found at https://www.paloaltonetworks.com/company/trademarks.html. All other marks mentioned herein may be trademarks of their respective companies.
cortex_eb_essential-guide-2023-mitre-engenuity-att&ck-evaluations_112123