The Essential Guide

to the 2024 MITRE ATT&CK Evaluations: Enterprise

paloalto NETWORKS | CORTEX XDR®

This e-book provides a comparative look at how vendors performed in the 2024 MITRE ATT&CK Evaluations: Enterprise across numerous measures, 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 to help you assess for yourself which vendor best fits your organization's endpoint security needs.

Since 2018, the MITRE 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 behaviors, the Evaluations provide the opportunity to analyze endpoint detection and response (EDR) products against real-world attack scenarios.

Table of Contents

What's Different This Year?

This year's Evaluations were centered around two themes: ransomware and emulation of attacks from North Korea, officially the Democratic People's Republic of Korea (DPRK), threat groups. Notably, this year's Evaluations expanded platform coverage and introduced testing for false positive detections and preventions.

These changes significantly increased the complexity of the evaluation, and may have contributed to a drop in participation of 10 *fewer* vendors this year. As threat actors continue to evolve their techniques, vendors face growing challenges in developing and maintaining security solutions that can effectively detect and prevent these sophisticated attack scenarios.

The Evaluations' structure has evolved to include:

- Multiple smaller emulations targeting specific threats
- Ransomware testing on Windows/Linux systems
- DPRK-attributed threat testing on macOS
- False positive testing scenarios
- Measuring alert volume

MITRE Adds False Positive Testing

The addition of false positive rates underscores MITRE's commitment to real-world applicability, as highlighted by William Booth, general manager of the evaluation:

> *This round will feature new insights, with a particular focus on efficiency, including true positive and false positive rates, which more accurately reflect the real-world performance of a tool.*
>
> **– William Booth, GM, MITRE**

The assessment now includes expanded scoring for detection and prevention accuracy, measuring both a solution's ability to identify genuine threats and its precision in avoiding false positives. This helps validate, for example, that security solutions can effectively block malicious activity while maintaining normal business operations.

With this change, MITRE now assesses how well vendors can maintain detection accuracy without defaulting to overly aggressive configurations. Adding false positives and alert volume to the methodology better simulates real-world challenges, where security teams need to identify sophisticated threats that blend in with normal business activities while avoiding alert fatigue.

In addition, MITRE eliminated the telemetry detection category from the 2024 Evaluations to prioritize actionable detections over raw data visibility. While telemetry coverage demonstrates a security tool's visibility across attack vectors, MITRE recognized that visibility alone doesn't necessarily translate to useful threat detection capabilities.

5 Years of Top Endpoint Security Results

Cortex XDR® has consistently ranked among the top performers in MITRE ATT&CK Evaluations, achieving leading detection results in every evaluation. This proven track record includes outstanding results against emulated adversaries and tools including Carbanak+FIN7 (2020), Wizard Spider and Sandworm (2022), Turla (2023), and DPRK, CL0P, and LockBit (2024), demonstrating robust protection against real-world threats.

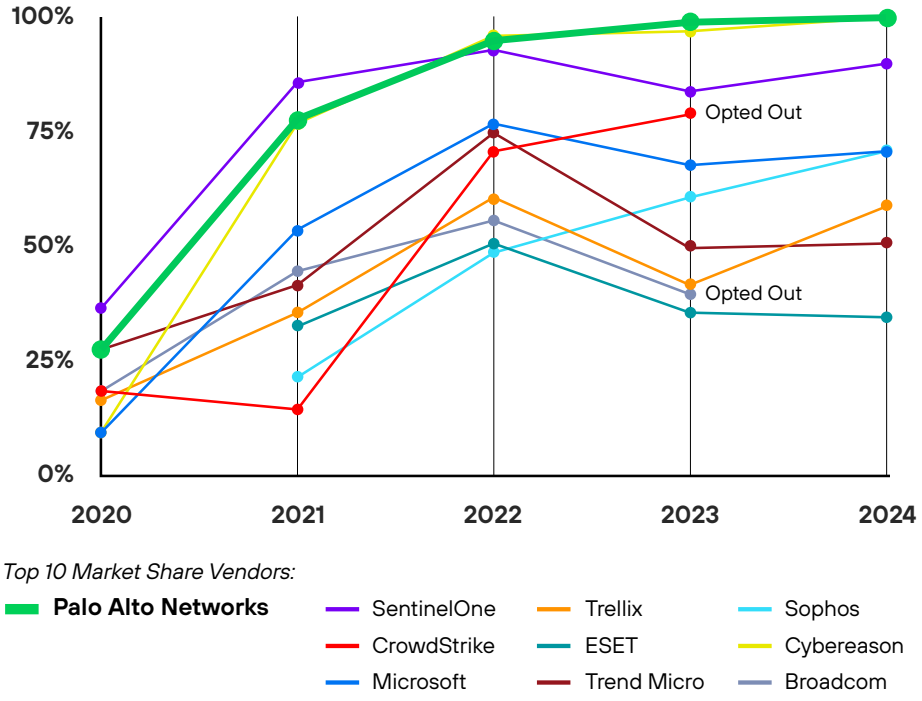

Figure 1: 5 years of technique-level detection rates

Cortex XDR Achieves 100% Detection with Technique-Level Detail and No Configuration Changes or Delays

For two consecutive years, Cortex XDR has delivered 100% detection coverage without requiring configuration changes or delayed detections. In the 2024 Evaluations, Palo Alto Networks was the first vendor ever to detect every stage of the three simulated attack scenarios with technique-level detail. Identifying attacks at the technique level provides security teams with the most detailed and actionable intelligence possible.

Overall, Palo Alto Networks rose to the challenge, delivering industry-best results:

- First-ever vendor to achieve 100% technique-level detection coverage with no delays or configuration changes.

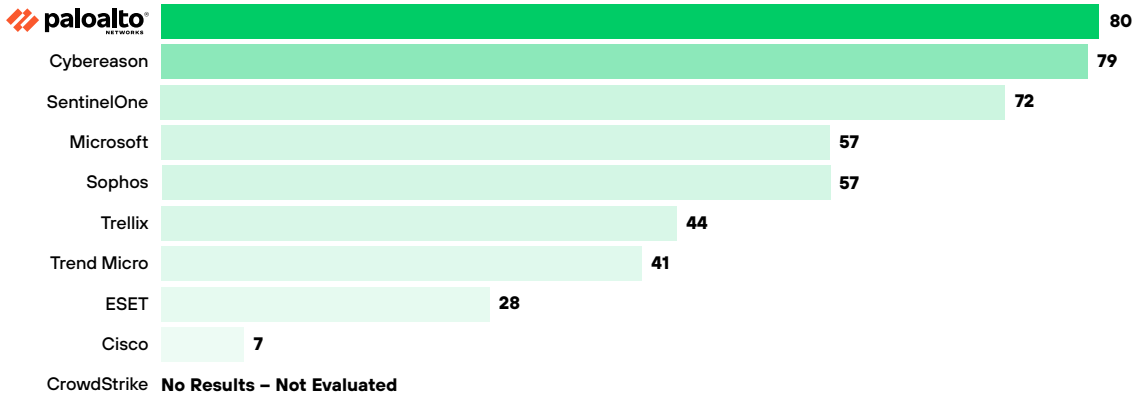

Figure 2: Technique detection coverage for the top 10 market share vendors

- 100% detection coverage across the expanded macOS and Linux attack surfaces.
- Highest prevention rate of all vendors with zero false positives that could disrupt critical business operations.

In addition to industry-leading detection rates, Cortex XDR blocked 8 of the 9 assessed attack steps in the prevention test—the best rate of all vendors that didn't have prevention false positives. Zero false positives in prevention is a critical distinction, as a single incorrect prevention event can interrupt normal business operations.

While there were two attack steps in the prevention test that didn't meet the criteria for MITRE to count as a block, the action Cortex XDR took during these steps would have protected customers and stopped a breach.

In step 3, the attack attempted an SSH connection from a suspicious host in China, which was supposed to be followed by MITRE's attack step. Cortex XDR blocked the suspicious SSH connection, stopping the attack at an earlier stage. As the intended techniques were not able to execute, this step was marked as Not Applicable, dropping the total assessed attack steps from 10 to 9.

In step 5, the attack attempted to encrypt data, and the encryption action was immediately reversed by the Cortex XDR agent. The attack was stopped, delivering the only outcome that matters, which is keeping customers safe.

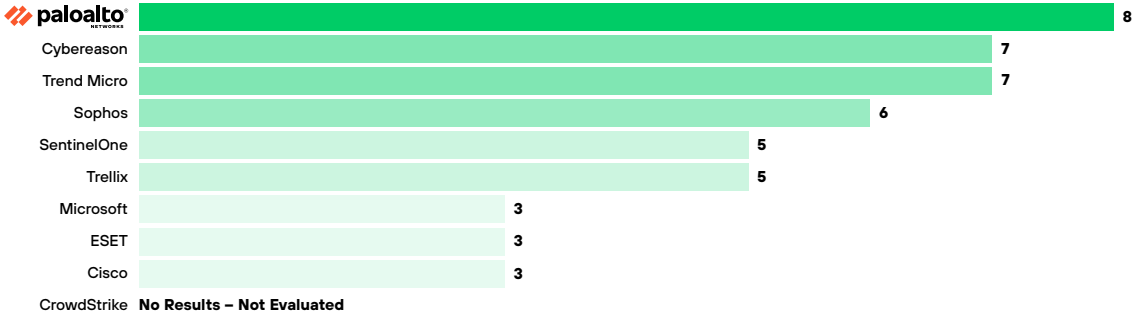

Figure 3: Attack steps prevented by the top 10 market share vendors

Evaluations Overview

The Adversary Scenarios

2024's MITRE ATT&CK Enterprise Evaluations introduce a targeted approach with two distinct focus areas: ransomware attacks on Windows/ Linux systems and DPRK-attributed threats targeting macOS environments. This shift from previous years' single large-scale to multiple smaller emulations allows for more nuanced assessment of defensive capabilities.

The choice of adversary scenarios reflects current critical threats:

- DPRK evaluation focuses on North Korea's expanding cyber operations targeting macOS systems, particularly relevant given their pattern of targeting high-value systems to fund nuclear capabilities.

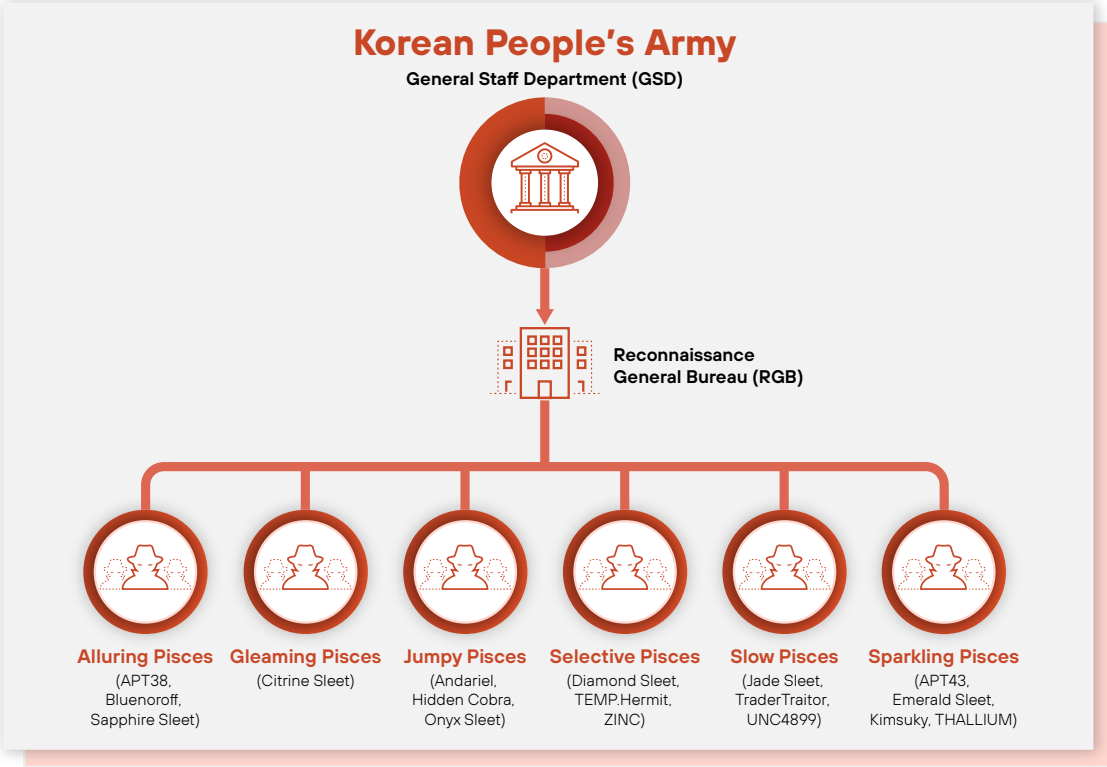

Figure 4: The DPRK has emerged as a formidable threat, progressively leveraging cyberattacks to fund the advancement of their nuclear capabilities

- Ransomware evaluation addresses the evolving ransomware-as-a-service (RaaS) landscape, where lower barriers to entry have led to increased global attacks.

This methodology provides a focused examination of defensive tools against these specific, highly relevant threat actors, offering organizations insights into security product performance against current real-world threats.

4 Evaluation Scenarios

1. DPRK on macOS
2. CLOP on Windows
3. LockBit on Linux
4. Protection evaluation

Detection:
16 steps and 80 substeps
(plus 20 false positives substeps)

Protection:
10 steps and 21 substeps
(plus 28 false positives substeps)

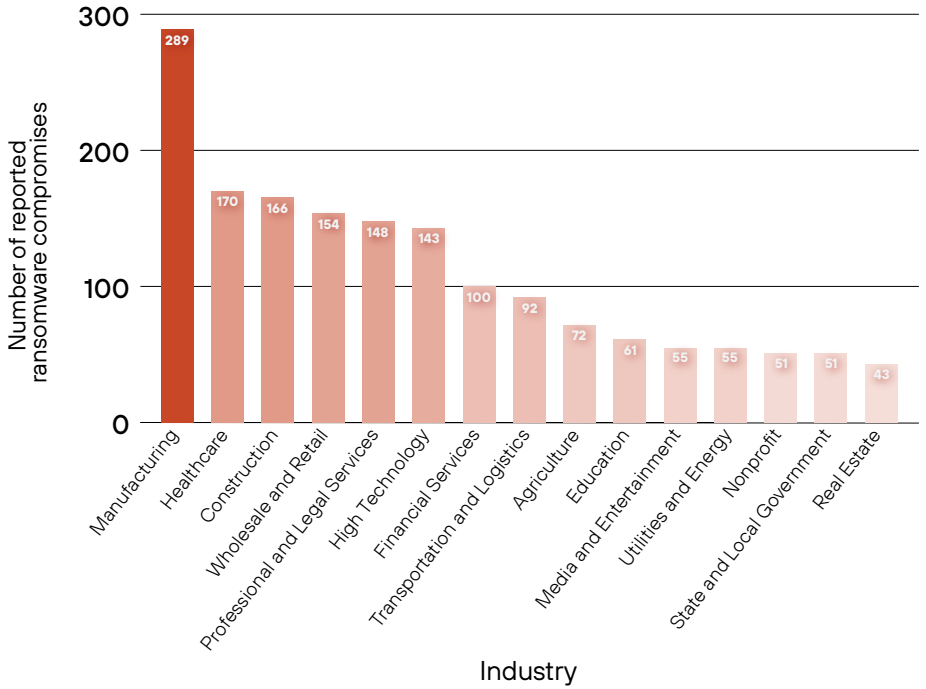

Figure 5: Ransomware continues to be one of the most significant global cybercriminal threats across industry verticals[1]

1. Kristopher Bleich and Amanda Tanner, "Ransomware Review: First Half of 2024," Palo Alto Networks, August 9, 2024.

Emulation Notes

Ransomware

These emulations explore common behaviors that are prevalent across prolific ransomware campaigns, such as the abuse of legitimate tools, data encryption, and disabling critical services or processes.

Democratic People's Republic of Korea

The macOS emulation features adversary behavior inspired by North Korea attacks on macOS via multistaged and modular malware in operations involving elevated privileges and credential targeting.

Technique Scope

The highlighted ATT&CK techniques in each diagram below are in scope for this evaluation. For Ransomware, the techniques are split amongst Windows and Linux. For DPRK, the techniques all target macOS.

Emulating Real-World Threats in the Cloud

The MITRE ATT&CK Evaluations' focus on North Korean threat groups and ransomware TTPs parallels the rise of cyberattacks targeting the cloud today. These threat actors employ sophisticated methods that are difficult to detect in distributed cloud environments, making this assessment useful for assessing the capabilities of cloud detection and response.

Highlights for Cloud Detection and Response (CDR):

Cloud environment testing: The Evaluations use Linux and Windows Server instances, simulating operating systems commonly run in the public or private cloud.

Full attack lifecycle visibility: Demonstrates complete visibility across the MITRE ATT&CK framework, essential for detecting lateral movement, privilege escalation, and data exfiltration in complex cloud environments.

Advanced ransomware defense: Evaluates the ability to identify early-stage ransomware behaviors targeting cloud-stored data.

Ransomware-as-a-service attacks against Windows and Linux systems. These emulations showcased common features across high-profile ransomware campaigns, like abusing legitimate tools, encrypting data, and disabling critical services or processes.

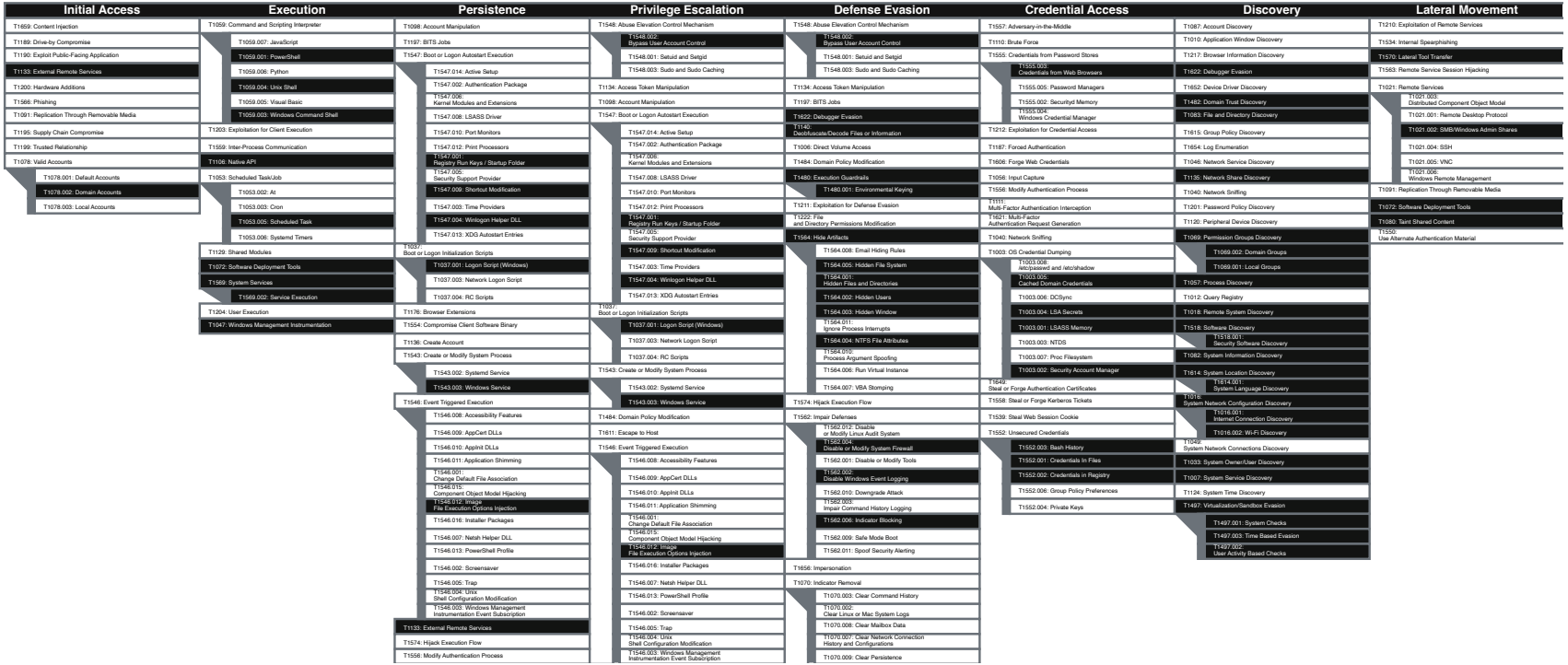

Figure 6: The MITRE ATT&CK framework—Ransomware, Windows and Linux

Attack patterns demonstrated by the DPRK against macOS. These emulations mimicked multistaged and modular malware in operations involving elevated privileges and credential targeting.

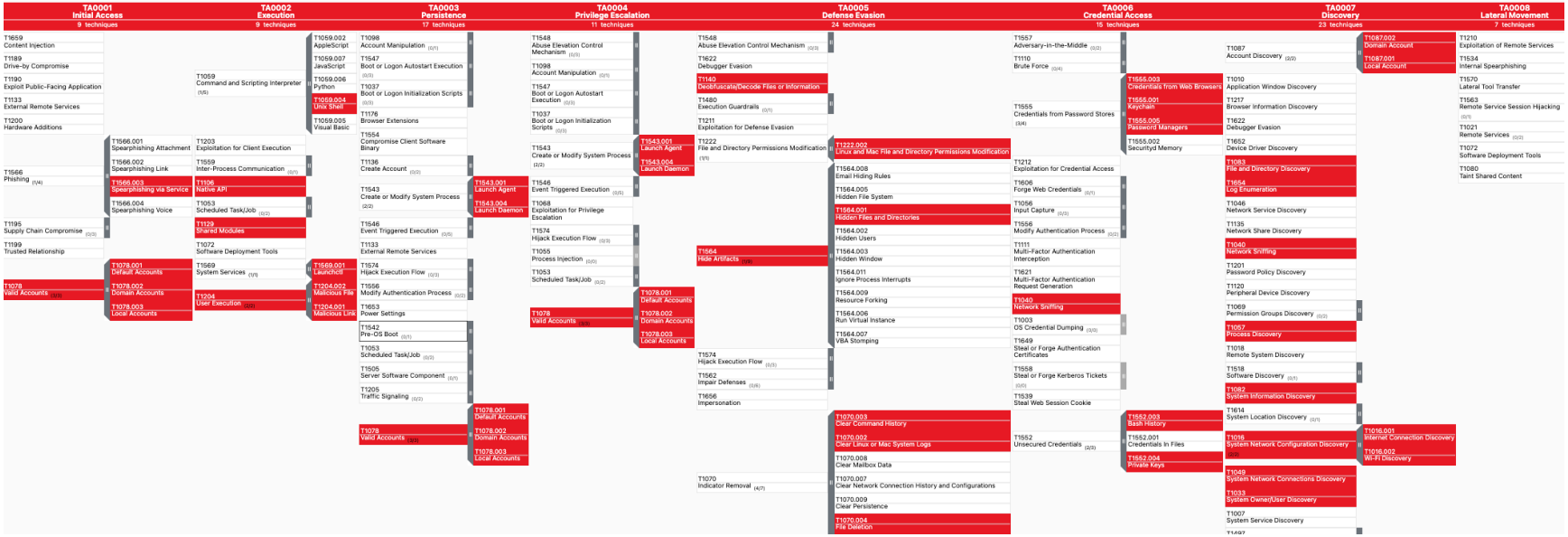

Figure 7: The MITRE ATT&CK framework—DPRK, macOS

The MITRE ATT&CK Framework

According to MITRE:

- *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.*

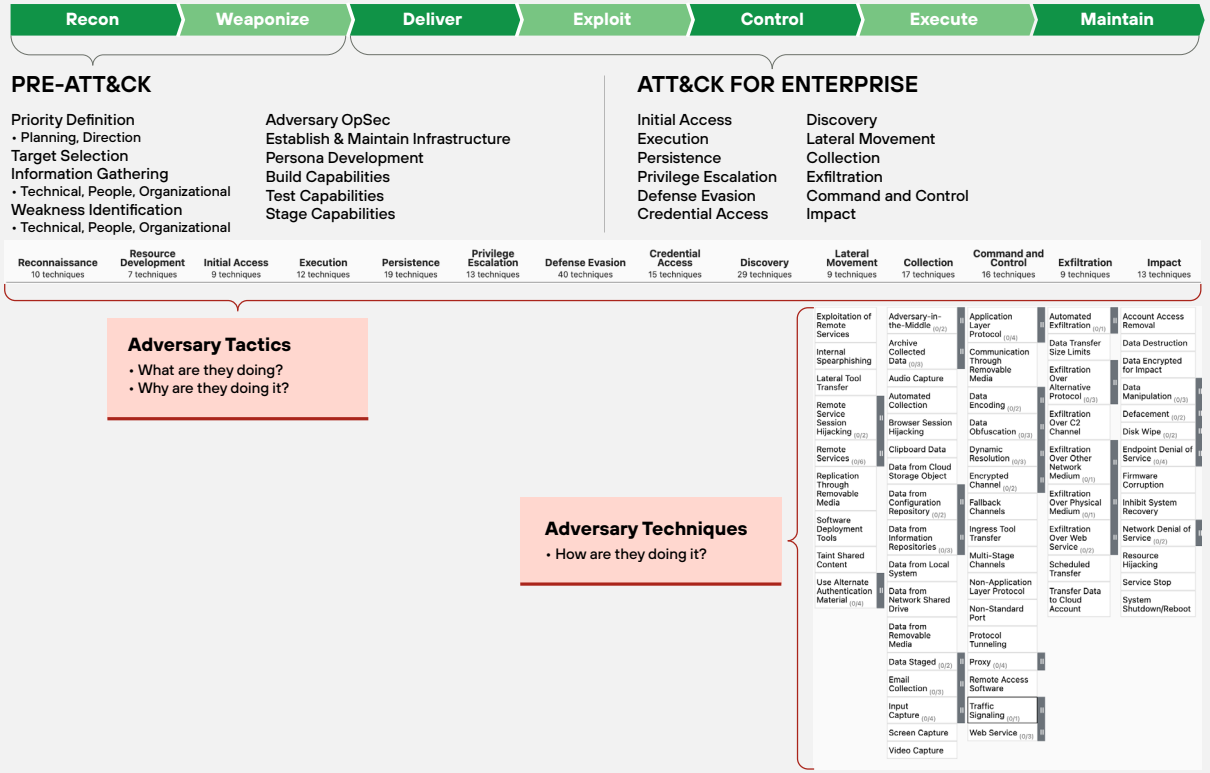

Figure 8: Understanding the MITRE ATT&CK framework

MITRE's Approach

MITRE categorizes each detection based on quality and precision, focusing on how detections occur rather than assigning scores. (See Detection Categories for more details.) While MITRE strives to capture detections independently, vendors might detect in ways that MITRE doesn't capture.

A detection must specifically apply to a given attack technique to be counted for that technique. For example, if a detection applies to one technique in a step or substep, it doesn't necessarily apply to all techniques in that step. MITRE requires vendors to provide proof of detection at each step and substep. Not all detection details may be included in public results, especially if they contain sensitive information.

To categorize a detection accurately, MITRE evaluates various materials and processes. These include reviewing screenshots of each detection, notes from the evaluation, responses to follow-up questions posed to vendors, and vendor feedback on preliminary results.

Additionally, MITRE conducts independent testing of procedures in a separate lab environment and examines detections from open-source tools and forensic artifacts. This comprehensive testing helps define what qualifies as a detection for each technique.

Using MITRE to Help Evaluate Endpoint Security Solutions

For organizations reviewing EDR solutions and vendors, the MITRE Evaluations 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.

> *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, ATT&CK Evaluations**

By using modern attack TTPs, solution providers can assess their performance and determine areas for improvement. The resulting performance data can provide insights into solution or product modifications that help improve security outcomes for customers.

MITRE Engenuity Round 6 Methodology

Detection Categories

The italicized category content that follows was taken directly from MITRE.

For this evaluation, there are 5 main detection categories representing the number of alerting context participants provide to the end user:

Technique

The solution autonomously identified that malicious/suspicious event(s) which meet the documented Detection Criteria for Tactic, and the collected evidence provides details to how the action was performed (technique) through the ATT&CK (sub-) technique or equivalent level of enrichment to the data collected pertaining to the behavior under test (the who, what, when, why, and where).

Tactic

The solution autonomously identified that malicious/suspicious event(s) which meet the documented Detection Criteria for General and the collected evidence provides details as to why the action was performed (tactic) through the ATT&CK tactic or equivalent level of enrichment to the data collected pertaining to the behavior under test.

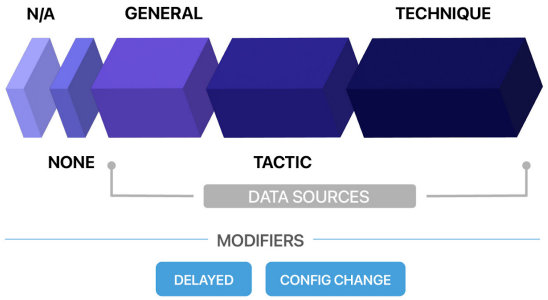

Figure 9: MITRE Round 6 detection categories

General

The solution autonomously identified that malicious/suspicious event(s) which meet the documented detection criteria occurred, and the collected evidence did not provide details as to why the action was performed (tactic) or how the action was performed (technique) through the ATT&CK tactic/technique or equivalent level of enrichment to the data collected pertaining to the behavior under test.

None

Execution of the sub step was successful; however, evidence provided by the vendor does not meet the documented detection criteria, or there was no evidence of Red Team activity provided. There are no modifiers, notes, or screenshots included with a None.

Not Applicable (N/A)

Evaluation for the (sub) step was not completed. Reasons may include, but are not limited to, technical issues during execution prevented successful completion, or the platform not supported by the participant.

Modifier: 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.*

Modifier: Delayed Detection

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 produce 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.

Learn More: Detection Categories

Learn More: Protection Categories

Additional Metrics

All italicized metrics content that follows was taken directly from MITRE.

False Positive

Benign activity (outside the emulation plan) will be executed during the evaluation while malicious/suspicious activity is executed. If the tool incorrectly alerts on or blocks benign activity as malicious or suspicious (via a detection or automated protection), it will be marked as a false positive.

Alerts must meet the general detection criteria, at a minimum, to be considered for this metric.

To determine a false positive rate, we will evaluate a subset of benign activity and determine the prevalence of false positives amongst the set under evaluation.

Examples:

- ***Legitimate File Sharing**: Benign file sharing activities between authorized users could trigger false alerts. Identifying this activity with a general alert would be incorrect and categorized as a false positive.*
- ***Background System Activities**: Innocuous background processes or system maintenance tasks. Identifying this activity with a general alert would be incorrect and categorized as a false positive.*

Alert Volume

Alert fatigue is a common challenge faced by security teams. Too many alerts can overwhelm analysts, leading to missed incidents or slow responses.

In the 2024 Evaluations, alert volume is shown as the raw number of alerts generated during the full evaluation period. Alert volume is recorded separately for the initial execution and the postconfiguration change execution.

Environment

Each evaluation is performed in a cloud computing environment with one or more victim organizations. The 2024 Evaluations contained the following operating systems:

- *Windows*
  - *Windows Server 2022*
  - *Windows 11*
- *Linux*
  - *Ubuntu 22.04.x LTS*
- *macOS*
  - *OS: macOS Sonoma 14.x*
  - *Arch: Apple Silicon*

Conclusion

As we reflect on these Evaluations, it's clear they provide valuable insights into the capabilities of endpoint 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 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 ATT&CK Evaluations Dashboard. As we continue to adapt to the ever-changing threat landscape, Palo Alto Networks remains dedicated to helping our customers defend against increasingly complex cyberattacks.

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:

- Visit our page dedicated to the MITRE Evaluations
- Watch our LinkedIn Live: MITRE Round 6: Results for the Toughest Evaluation Yet
- Read our blogs:
  - MITRE ATT&CK 2024: Raising the Bar for Security Testing
  - Cortex XDR Delivers Unmatched 100% Detection in MITRE Evals 2024
  - MITRE ATT&CK Evaluations — Cortex XDR Among Elite in Endpoint Security

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 don't 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.

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_mitre-att&ck-round-6_022825