The Ultimate Guide to the MITRE ATT&CK Round 2 EDR Evaluation

Marketing jargon can make cybersecurity products hard to differentiate, so third-party testing that compares products head-to-head is invaluable. Starting in 2019, the MITRE ATT&CK® cybersecurity evaluations have quickly become counted among the most useful objective tests for endpoint detection and response (EDR) solutions, providing a wealth of information about the efficacy of tools by methodically testing their detection and correlation capabilities against the attack sequences of real-world adversaries. The second round of MITRE ATT&CK testing, released in April 2020, assessed a much wider field of vendors than the first, giving security decision-makers a comprehensive view of how endpoint security players stack up.

MITRE publishes raw data based on in-depth testing, giving vendors and analyst communities the ability to create scoring systems that synthesize the results and help buyers make decisions. This guide provides a comparative look at how vendors performed across multiple measures, with guidance on how you can explore the results further. We'll walk you through MITRE's testing methodology, the tools MITRE provides to help you visualize and compare results, and various considerations for analysis to help you assess for yourself which vendor best fits your organization's endpoint security needs.

MITRE Round 2 Methodology

Round 2 was designed to emulate the APT29 threat group, a.k.a. Cozy Bear—a sophisticated adversary group associated with nation-state activity. APT29 is known for stealthy attacks that utilize an arsenal of custom malware and varied operational cadences. In this round, MITRE created two different attack scenarios: one that emulated a "smash-and-grab" attack, and one that was much more targeted and deliberate.

For each of the 58 attack techniques tested, MITRE documented whether each vendor product detected that technique and the type of detection (or detections, as each step could have more than one), on a scale ranging from no detection (labeled as "None") up to alerts with information about the specific technique used ("Technique"). MITRE also captured whether human monitoring and analysis played a role in a detection by applying the "MSSP" label.

Figure 1: MITRE evaluation detection categories

On top of these categorizations, MITRE applied modifiers as necessary:

- **Alert:** The detection generated a notification for analysts. This only applies to the General, Tactic, or Technique detections.
- **Correlated:** Data was tagged to show an association with a previously discovered malicious/suspicious behavior. Known as "Tainted" in last year's test, this modifier was renamed to make it clear that it is a positive modifier.
- **Delayed:** The detection did not occur in real time; it was delayed either because it was found by a human (such as with an MSSP) or after additional complex data processing.
- **Host Interrogation:** Data did not trigger an automatic detection but was available to be pulled manually from an endpoint during analysis. This is only useful to experienced analysts who may manually discover the data during in-depth investigations as it isn't implemented programmatically, which is why MITRE doesn't count it as a detection.
- **Residual Artifact:** Data did not trigger an automatic detection, but it can be manually pulled and analyzed to determine that certain attack capabilities or behaviors were used. Again, this is only useful to experienced analysts during in-depth investigations as it isn't implemented programmatically, which is why MITRE doesn't count it as a detection.
- **Configuration Change:**
  - **UX Changes:** A change was made to the product during the course of the test that affected the user experience but not the detection capability.
  - **Detection Changes:** A change was made to the product during the course of the test that enabled a detection that otherwise would have been missed.

How MITRE Can Help You Evaluate EDR Solutions

So, you're in the market for EDR. How do the MITRE ATT&CK results help you pick the right tool for your organization? While you may weigh components differently according to your needs, MITRE's data contains several key measures of efficacy that are universally relevant to security teams. These measures should be considered holistically, rather than in isolation. They include:

- **Overall detection**, which assesses a solution's ability to detect a threat at all.
- **Correlation and quality of detections**, which describe the levels of information that each detection provides analysts.
- **Actionability of detections**, which accounts for the other measures, but also factors in how quickly analysts can act on information. Can the product group alerts into incidents and provide root cause analysis? How well does it generally support security teams' workflows? Actionability is quite subjective as it accounts for needs and preferences regarding things like user interface (UI) and use of managed services.

Overall Detection Capabilities

This is the first metric to consider for each vendor. Whatever the quality of each detection, some form of detection must occur for an analyst to be able to investigate and respond. Figure 2 shows the percentage of techniques for which each vendor's product had any kind of detection, excluding those tagged with the "Configuration Change" modifier, as product configurations aren't made mid-attack in real-life attack scenarios. In the APT29 test, vendor detection scores ranged from 47% to 90%. We're proud to share that Cortex XDR™ by Palo Alto Networks was unsurpassed in overall visibility.

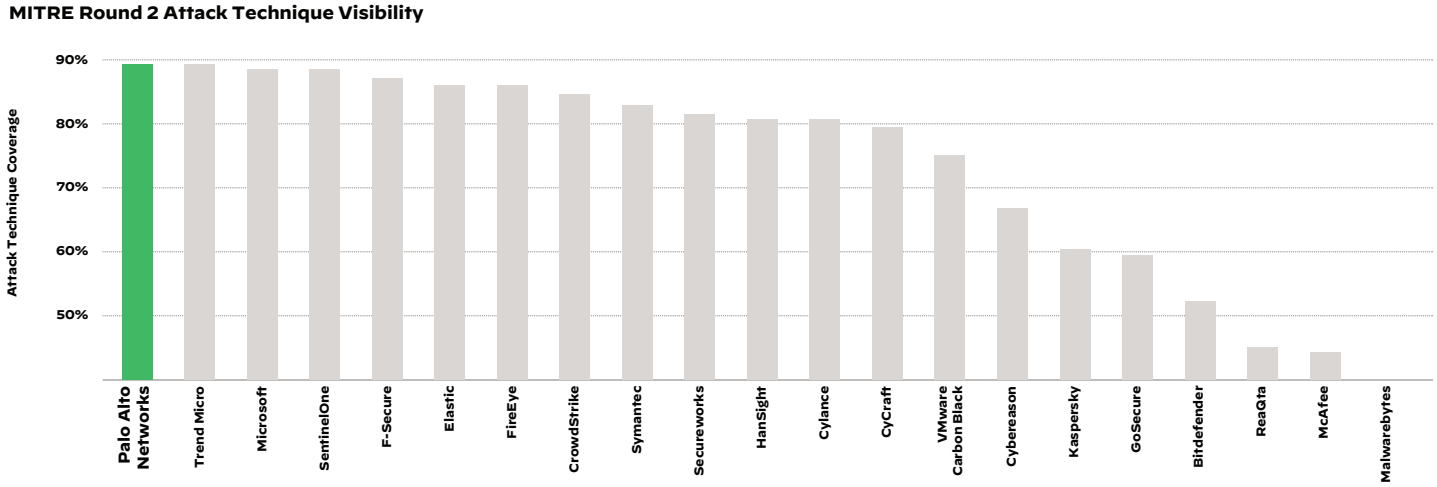

Figure 2: Overall attack technique coverage

Detection Quality

Of course, detection quality matters. Enterprises receive an average of more than 11,000 security alerts per day, many of them false positives. Security teams don't need more alerts—they need better ones. Security solutions should maximize the fidelity of the alerts they deliver, the information contained within each alert, and the enrichment of those alerts by linking them to other correlated security events.

Tactics and Techniques

Using MITRE's taxonomies, "Tactic" detections (which include information about an attacker's intent, or why an activity may be happening) and "Technique" detections (which give information about both why and how it is happening) are the detection types that contain **the most information about that specific step in the attack**.

Figure 3 shows how many of these types of detections each vendor produced in the APT29 test. We've removed tactics and techniques that were flagged directly with the "Configuration Change" modifier, but configuration changes still may be skewing this data in cases where a configuration change led to the tool making a lesser detection (e.g., "Telemetry"), which the vendor's MSSP team then manually followed up on to generate a Tactic or Technique detection over the top of it.

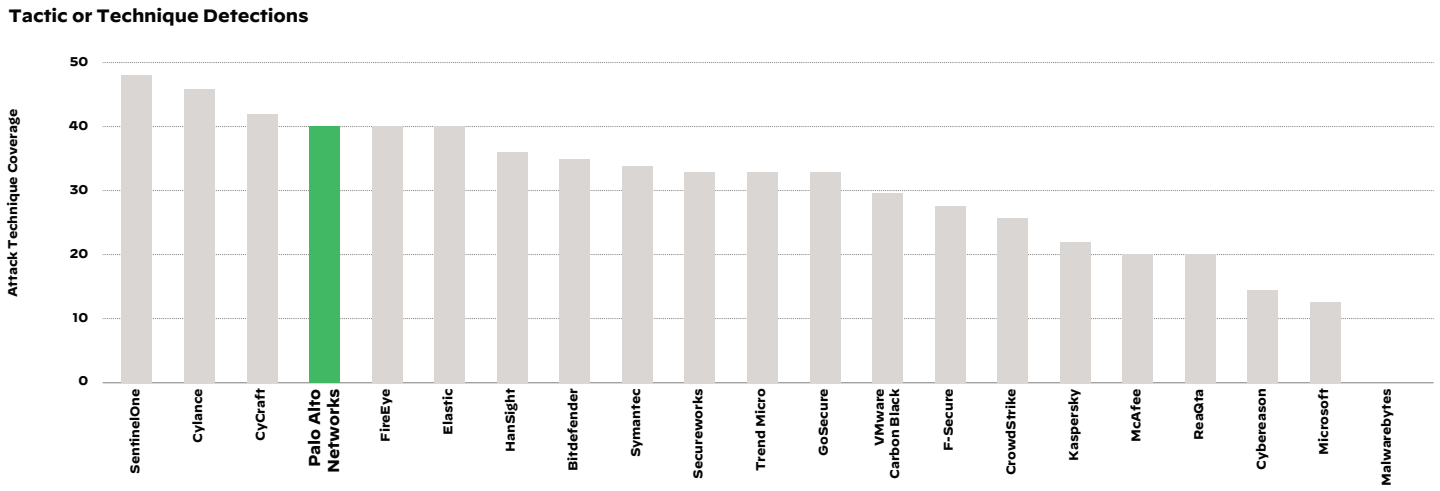

Figure 3: Sum of "Tactic" and "Technique" detections

Correlations

While "Tactic" and "Technique" detections provide the most information about a specific attack step, the "Correlated" modifier indicates whether that technique is linked to other attack steps, which is critical context for triage and investigation. If a detection is "Correlated," it's linked to other security events that happened during the same attack chain. Vendors whose products can find these linkages and deliver a higher percentage of "Correlated" detections provide superior security analytics that often result in higher alert fidelity in general. Correlations are key to how Cortex XDR groups events into incidents, reducing the number of disparate alerts analysts need to see by 98%. Figure 4 shows the number of correlated events per vendor.

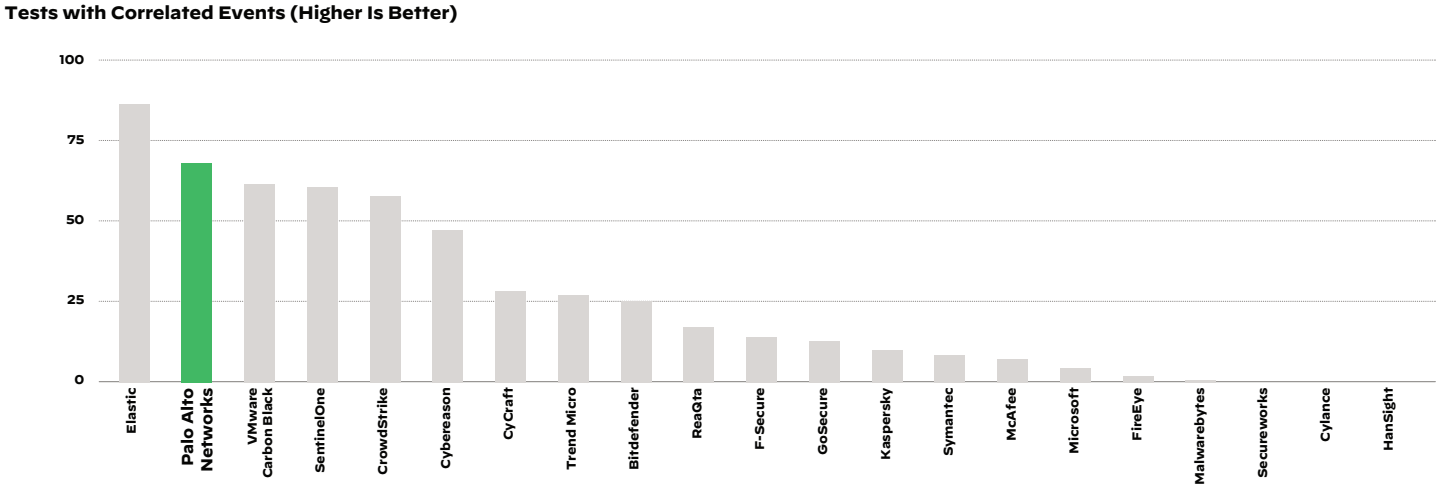

Figure 4: Number of tests with correlated events

Actionability

Figure 4 should give you a feel for how effective various solutions are when it comes to detecting and generating information about threat techniques. Still, detection is only helpful if a security team can act on it. In reality, security teams face many noisy alerts on a daily basis, so an EDR solution must not just add more alerts—it must help security teams make sense of the noise to quickly triage, investigate, and remediate confirmed threats. Therefore, security teams should consider the "Actionability" of each tool and/or service based on how well it can identify the highest priority security incidents and group related alerts together in a way that enables analysts to easily investigate and take response actions. MITRE's data contains useful information about both the UI of the tool as well as the capabilities of vendors' managed services to help you with this assessment.

User Interface

The UI matters quite a bit to your analysts' ability to use tools efficiently and effectively—as well as to your ability to onboard and train junior analysts. If you're willing to dive deeper into the data, MITRE offers a Technique Comparison Tool that lets you compare vendors head-to-head, including screenshots of each detection to show product usability. This is an easy way to understand what it actually feels like to use a tool, looking across different vendors to determine how it will work within your SecOps team's processes and workflows.

To use this tool, first select "Technique Comparison Tool" from the "Tools" drop-down in the menu bar of the MITRE website.

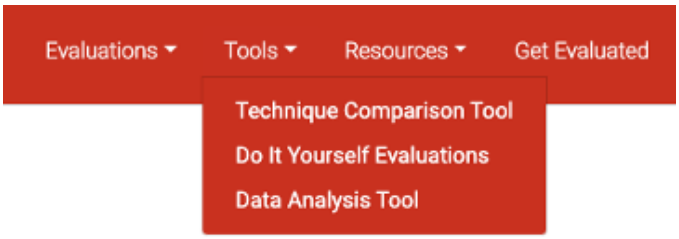

Figure 5: MITRE website drop-down menu

Second, select the vendors you'd like to compare, and then click through the various techniques to see the detections they delivered in each step of the MITRE ATT&CK Evaluation.

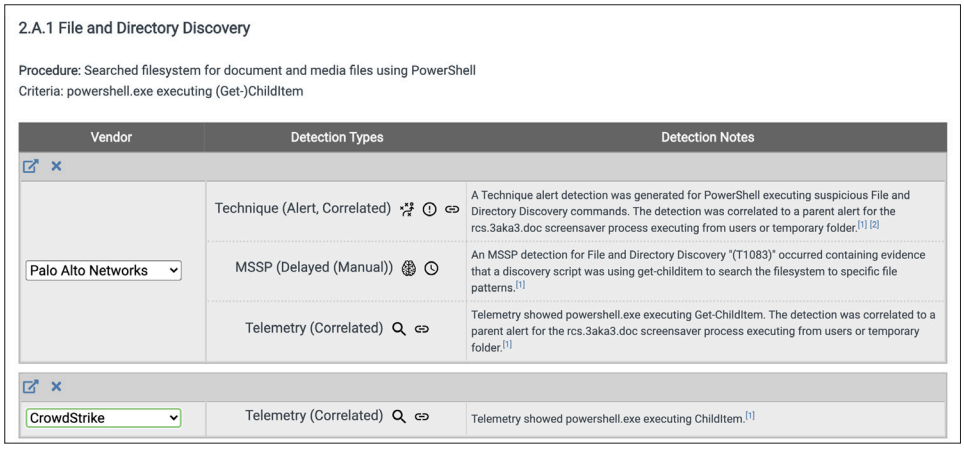

Figure 6: Technique comparison tool

Finally, click the superscript numbers in the Detection Notes to see screenshots of the detections. Figure 7 is one such screenshot showing Cortex XDR. As you can see, Cortex XDR integrates data to show the full flow of an attack within its UI, tying together all correlated alerts in a clear and easy-to-follow chain of events. Compare this to other vendor products and you'll see that the levels and presentation of information are noticeably different.

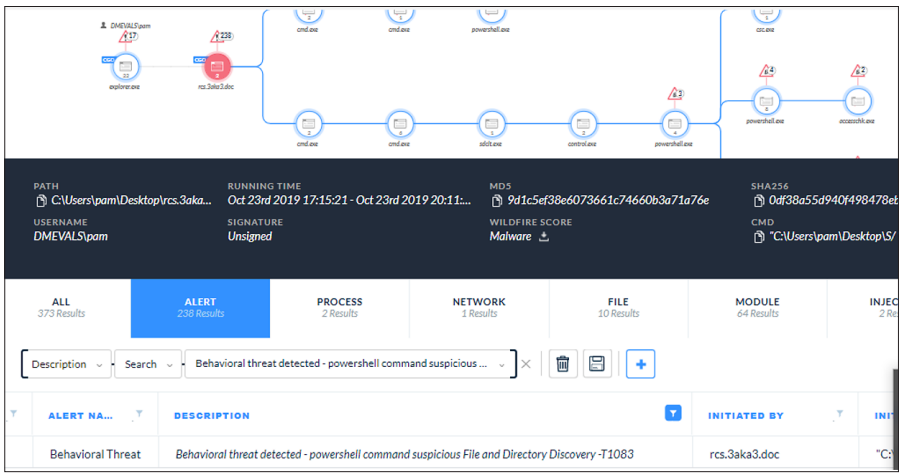

Figure 7: MITRE screenshot of Cortex XDR

Managed Services

SecOps teams are structured in all kinds of different ways. A recent study by Forrester Consulting showed that more than half of companies lack a formal security operations center (SOC), and even those that have one often outsource some functions, including detection, response, and threat hunting. In this test, detections with the "MSSP" label were generally delivered by the vendors' managed detection and response (MDR) or managed threat hunting (MTH) teams. If your organization doesn't have sufficient expert resources to dedicate to threat hunting and you're considering working with a partner, "MSSP" detections are worth taking a look at, as the MITRE evaluation revealed a wide variance in the efficacy of vendors' services. Adding up and/or comparing the real-time detections (which are delivered by the tool) and the MSSP detections (which are delivered by a service) will give you a feel for how different delivery models may fit with your organization's needs.

The Cortex XDR Managed Threat Hunting service was still in beta during the MITRE APT29 evaluation, yet it had a very strong showing with 100 detections (see figure 8), none of which were linked to any configuration changes.

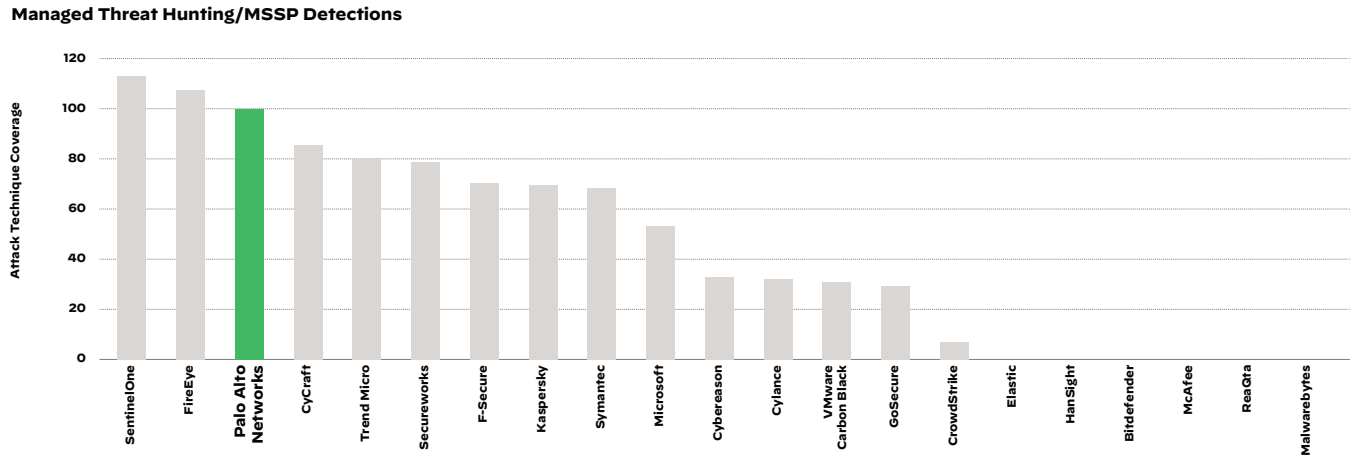

Figure 8: MSSP detections

What MITRE Doesn't Tell You

As much valuable information as the MITRE ATT&CK Evaluations contain, it's important to understand that the tests have a limited scope and should therefore be just one tool in your overall evaluation of any EDR vendor. As you weigh the findings against your organization's needs, make sure you consider a few things.

Product Tweaks for Testing Environments

Vendors are compelled to optimize their tools to perform well in each test. For a test like the MITRE ATT&CK Evaluation, which does not measure false positives, that means tuning their platforms to be extremely sensitive, often generating a high number of alerts that would completely overwhelm security teams in real production environments. Looking at quality and correlation of alerts can give a sense of the quality of each vendor's analytics, but in the end, MITRE only counted valid detections in this test, so you can't use the results to estimate a vendors' overall accuracy rates.

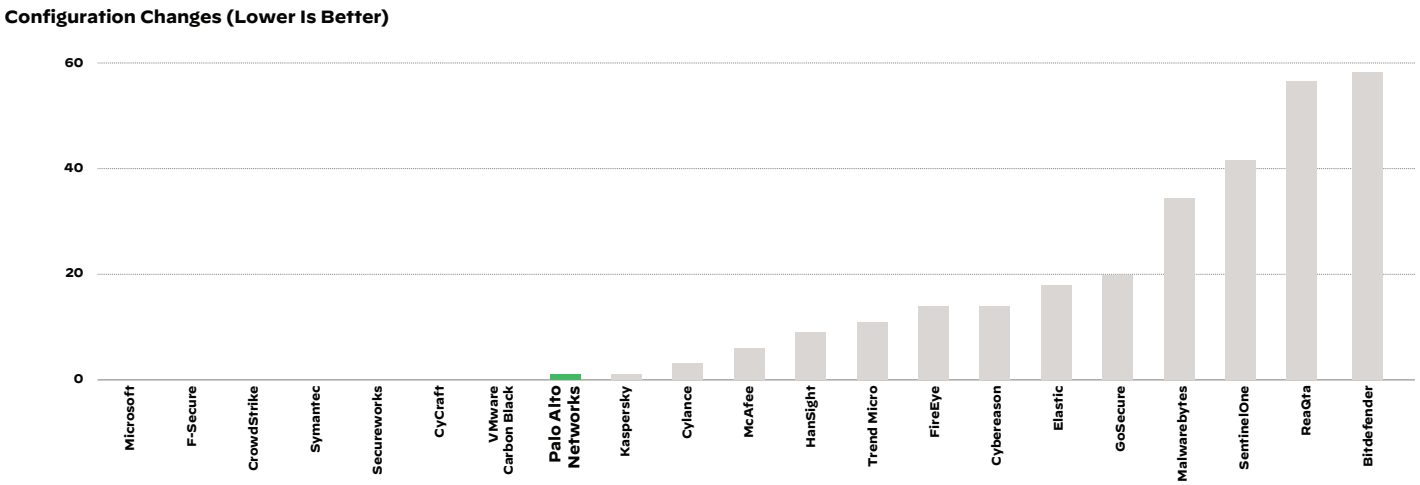

Figure 9: Configuration changes

Some vendors took their product tweaking a step further and made configuration changes mid-evaluation. While MITRE tells you how many such changes were made (a substantial number, in some cases), the results don't offer details about exactly what the changes were. When your security team uses a tool to detect real attacks, they won't have the benefit of these on-the-fly configuration changes, so you should take every detection flagged with a "Configuration Change" modifier with a grain of salt.

Endpoint Prevention Capabilities

The ultimate security outcome for any organization is to prevent attackers from ever entering your infrastructure. In the MITRE ATT&CK Evaluations, endpoint protection is turned off entirely—so while you get information on how well vendors can detect attack techniques, you don't get any information about how many of those attack techniques a vendors' endpoint protection capabilities would prevent outright. Other testing, such as the NSS Labs Advanced Endpoint Protection (AEP) Test, evaluates this component and should be a factor in any holistic evaluation.

Extended Detection and Response (XDR) Capabilities

XDR, or extended detection and response, describes detection and response systems that can ingest and analyze data from multiple sources, such as endpoint, network, and cloud. Gartner has listed XDR as one of the top security and risk management trends of 2020.

XDR is gaining traction rapidly among security operations teams not only because it helps consolidate the security technology stack and broaden visibility into the full scope of an attack, but also because it allows for better alerts. XDR can combine softer signals from multiple components to detect events that might otherwise be ignored. At the same time, XDR solutions can validate alerts by analyzing them with much broader context, resulting in higher fidelity. The net result of all this data stitching is that organizations have greater visibility, faster investigations, more comprehensive and more accurate alerts, and lots of good data feeding into their machine learning models to continue improving their analytics.

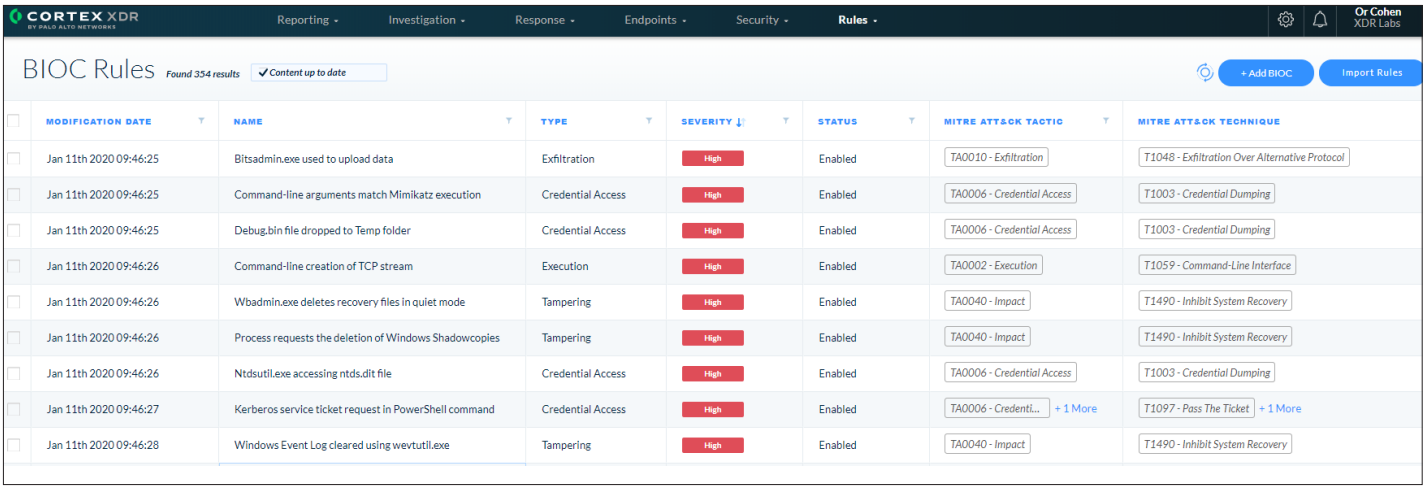

Figure 10: Cortex XDR stitches log data together to display attack scope and root cause

The MITRE ATT&CK APT29 Evaluation focuses on endpoint data, so for now, XDR capabilities must be evaluated separately. In the future as XDR maintains its momentum, MITRE may expand the testing methodology to include new types of security data.

The Cortex XDR Difference

We're extremely proud of the performance of Cortex XDR in the MITRE ATT&CK Evaluations. Cortex XDR was not only unsurpassed in overall detections; it also ranked highly across each of the aforementioned quality measures, showcasing its leading EDR capabilities.

Cortex XDR is much more than just a leading EDR (and endpoint protection) solution, however. It also delivers full-scale XDR capabilities across network, endpoint, and cloud data to help your organization scale and mature your security operations as well as stop sophisticated attacks.

Cortex XDR automatically stitches together different types of data and reveals the root cause of alerts, allowing analysts of all experience levels to perform alert triage and incident investigation in one console. Machine learning and AI models uncover threats from any data source, including managed and unmanaged devices. To further help your analysts understand attackers' methods

and objectives at each stage of an attack, Cortex XDR tags detections with the associated MITRE ATT&CK technique and tactic for every alert that relates to the MITRE ATT&CK framework. Finally, tight integration with enforcement points lets security teams respond to threats quickly and apply the knowledge gained from investigations to detect similar attacks in the future. Cortex XDR has reduced investigation times by 8x in our own SOC, and it can do the same in yours.

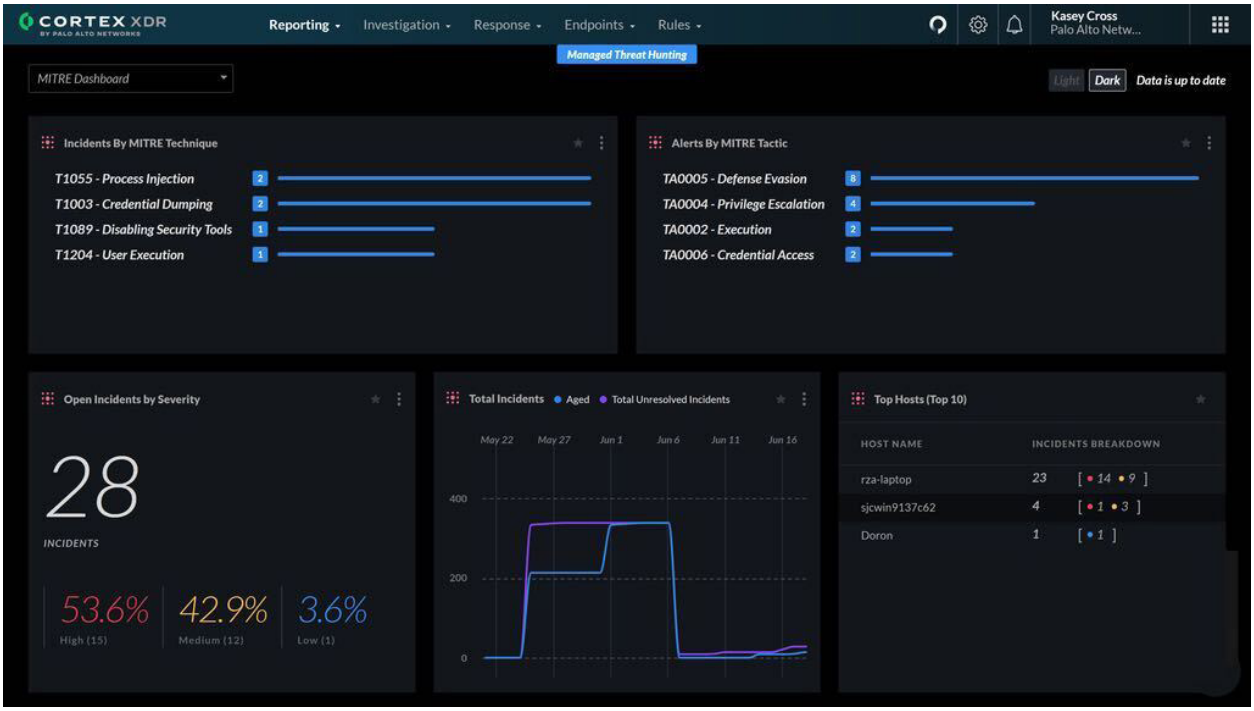

Figure 11: MITRE tagging in Cortex XDR

We offer additional support with the Cortex XDR Managed Threat Hunting service. Your security team can choose to augment the power of Cortex XDR with the expertise of our globally renowned Unit 42 threat intelligence team to identify hidden attacks that would otherwise go undetected. Our threat hunters combine years of knowledge and experience with big data analytics and comprehensive threat intelligence to surface malicious tactics, techniques, and procedures hiding among billions of benign actions.

To see more customer, analyst, and third-party testing validation of how Cortex XDR can help your security operations team deliver the best security outcomes for your organization, visit our website.

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