MITRE ATT&CK EVALUATIONS–APT29

Trend Micro's Results

ABOUT MITRE ATT&CK

MITRE ATT&CK is a public knowledge base of adversarial tactics and techniques, which can be used as a foundation for the development of specific cyber threat models and methodologies. In short, it helps the industry define and standardize how to describe an attacker's approach. MITRE ATT&CK collects and categorizes common attack tactics, techniques, and procedures (TTPs), then organizes this information into a framework. This framework can be used to help explain how adversaries behave, what they are trying to do, and how they are trying to do it.

Having a common language and framework is important in the ability to communicate, understand, and respond to threats as efficiently and effectively as possible. It also helps SOC/IR teams understand what coverage they have against various attack techniques. The framework is updated regularly with new techniques contributed by those in the cybersecurity industry, including Trend Micro. The MITRE ATT&CK evaluations have focused on the Enterprise Matrix for Windows systems, to-date, however, there are multiple framework matrices:

- Enterprise (Microsoft® Windows®, macOS®, Linux®)
- Cloud (Microsoft® 365®, AWS, Microsoft® Azure™, Google Cloud Platform™, Software as a Service (SaaS))
- Mobile (Android™, iOS)
- Industrial control systems (ICS)

Date

Evaluation results published April 21, 2020

Event

Third-party evaluation of a vendor's ability to detect adversary behaviors.

https://attackevals.mitre.org/evaluations.html?round=APT29

Trend Micro Solutions included in the evaluation

- Trend Micro Apex One™ as a Service with Endpoint Sensor (primary product)
- Trend Micro™ Deep Security™ enterprise
- Trend Micro™ Deep Discovery™ Inspector
- Trend Micro™ Managed XDR Service

TREND MICRO'S MITRE ATT&CK EVALUATION RESULTS

How the evaluation works:

MITRE ATT&CK provides a lab environment for vendors to install their products. Given the evaluation assesses post-compromise behavior, vendors must configure their products in detection or "alert only" mode. Then, a simulation of an advanced persistent threat (APT) attack is executed, and the vendor's solution is evaluated for what kind of techniques it is able to detect based on the MITRE ATT&CK Matrix for Enterprise. Each evaluation looks at a different attack scenario in the style of a real-world adversary group.

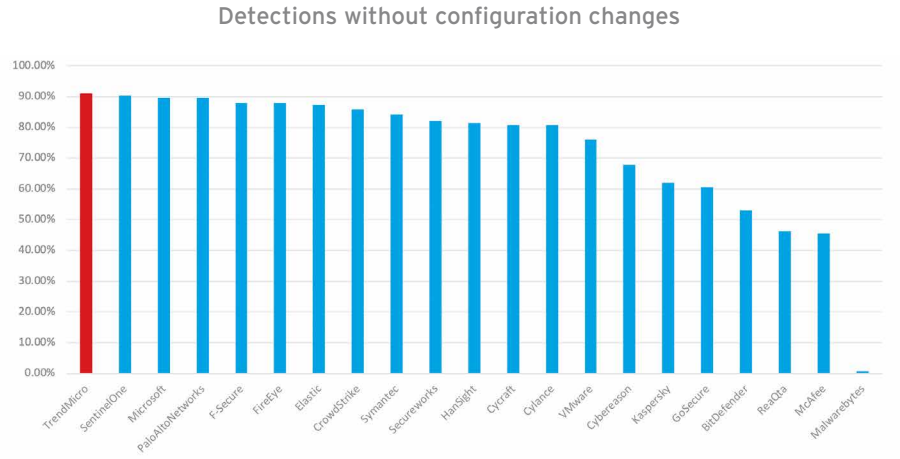
This evaluation emulated APT29-like behavior. APT29 (also known as Cozy Bear, The Dukes) is a threat group that has been attributed to the Russian government, operating since at least 2008. This group reportedly compromised the Democratic National Committee, starting in the summer of 2015. APT29 is distinguished by the stealth and sophisticated implementations of techniques via an arsenal of custom malware.

Overall results for Trend Micro

- **Trend Micro placed first in detections,** based on the initial product configurations—91% detection rate.
    - The evaluation allowed vendors to make product adjustments after a first run of the test to boost their detection rates on a re-test. MITRE ATT&CK results reflect vendor detection rates after all product adjustments, where we had the second highest detection rate coverage overall.
- **Lowest number of missed detections** among all vendors (for initial configuration).
- **Strong technique detections,** which is a higher confidence detection type.
- **Managed alert volumes**. A lower level of alerts combined with high detection rate means we can reduce the noise and avoid alert fatigue for overtaxed teams.
- **High amount of telemetry collected**—telemetry equals visibility. Trend Micro collected 103 pieces of telemetry, and on this measure, were among the top tier of vendors.
- **Managed Detection and Response (MDR) enriched detections**, however, detection coverage results would have remained strong without our MDR service.

FAQS

Can you explain detection rates?



This evaluation allowed vendors to make product adjustments after a first run of the test to boost their detection rates on a re-test. The final results shown reflect their detection rates after all product adjustments. If you assess what the product could detect as originally provided, we had the best detection coverage among the pool of 21 vendors. We believe this is a good way to consider the results for a few reasons:

- Once the initial test is done, vendors assess the results and then they can make front-end (UX) and back-end (detection) configuration changes and the test is run again. Product adjustments can vary in significance and may or may not be immediately available in vendors' current product.
- It is easier to do better once you know what the attacker is going to do. In the real world, customers don't get a second try against an attack.

Without making any kind of exclusions to the data, and just taking the MITRE ATT&CK results in their entirety, we had the second highest detection rate coverage. Considering our collective detections, versus the total number of steps evaluated, we had over 91.79% detection coverage and placed second in the pool of 21 vendors—showing a great balance of detection capabilities across the attack chain.

Additional considerations

- Automated detection and response.
    - This evaluation only tests detection after an event. In at least 10 steps of this targeted attack, Trend Micro's automated detection and response would have intervened and interrupted the attack with a blocking action (kill process, quarantine, isolate, etc.)
- Trend Micro™ XDR platform was not part of this evaluation
    - Correlation and context are the primary areas of focus for XDR

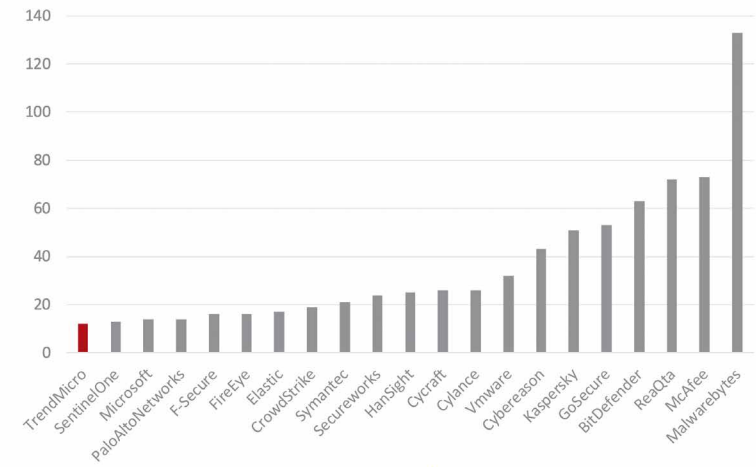
Full results here:
https://attackevals.mitre.org/APT29/results/trendmicro/

What happened with the 11 detections we missed (listed as "None")?

Detection misses are on the other side of the detection coverage equation. Given we had great coverage, we had few misses. We did extremely well against the competition, in terms of the number of missed detections. We had the lowest number of missed detections (12) among all vendors, based on initial product configuration and had the second lowest (11) in the final results, after product adjustments.

Remember that these detections are not individual attacks, they are small steps in a larger attack. It is not necessary to detect every single step in the attack in order to detect it and respond appropriately. Regardless, this test enables us to identify areas to improve the product and some items related to missed detections are now in the development queue as a result.

Fewest Missed Detections — Initial Configuration



What does configuration change mean?

In these cases, something might not have been detected, but the product was capable of detecting it with a configuration change. The vendor made a product adjustment and the detection was made on a second attempt.

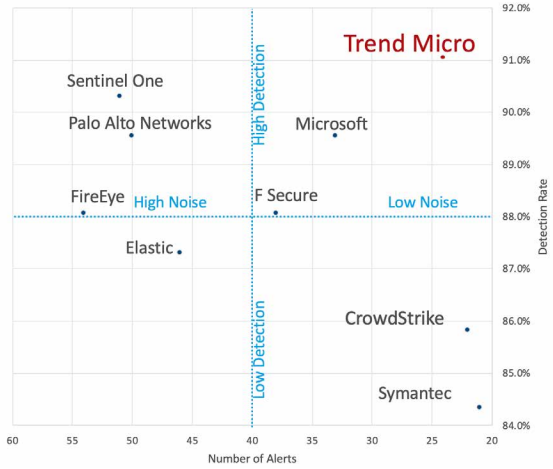
As noted above, configuration changes did not have a material impact on our overall detection coverage results. In fact, when you compare vendors, removing detection results post-configuration changes, Trend Micro placed first for initial overall detection. This is another area where the configuration changes made during the test have helped us identify and initiate product improvements.

Trend Micro's number of alerts are lower: Why is that a good thing?

At first glance, some may expect we should have the same number of alerts as detections. However, not all detections are created equal and not everything should have an alert. Too many alerts can lead to alert fatigue and add to the difficulty of sorting through the noise to get to what is most important. For example, one of the vendors in the evaluation triggered 90 individual alerts from the APT attack, creating an excessive number of items for a security analyst to triage.

In comparison, Trend Micro had a third of those alerts, making it more manageable for security to review and get to what was important. When you consider the alerts associated with our higher-fidelity detections, Trend Micro did very well at reducing the noise of all of the detections into a minimal number of meaningful alerts.

Highest Initial Detection, Low Alert Volume
(Selected vendors shown)



What is the difference between "General", "Tactic", and "Techniques"? Trend Micro seemed low on General and had no Tactic detections, is that a problem?

https://attackevals.mitre.org/APT29/detection-categories.html

There is a natural hierarchy in the value of the different types of detections:

- A General detection indicates that something was deemed suspicious, but it was not assigned to a specific tactic or technique.
- A detection on Tactic means the detection can be attributed to a tactical goal (e.g. credential access).
- A detection on Technique means the detection can be attributed to a specific adversarial action (e.g. credential dumping).

We have strong detection on Techniques, which is a better detection measure. With the individual MITRE ATT&CK Technique identified, the associated Tactic can be determined, as typically there are only a handful of Tactics that would apply to a specific technique. When comparing results, you can see that many vendors had lower Tactic detections on the whole, demonstrating a general acknowledgement of where the priority should lie. The fact that we had lower General detections compared to Technique detections is a positive.

In the initial run of the test, we collected 103 pieces of telemetry, making us among the top tier of vendors (range for telemetry collected among vendors was 1 to 113). This demonstrates that we give security analysts access to the type and depth of visibility they need when looking into detailed attacker activity.

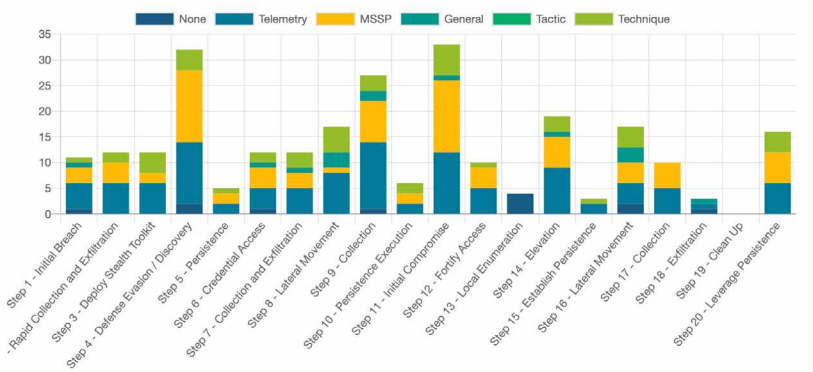
How much did MDR influence the results?

Trend Micro MDR analysts contributed to the "delayed detection" or MSSP category. This is where the detection involved human action and may not have been initiated automatically.

Our results show the strength of our MDR analysts. If an MDR service was included in this evaluation, you would want to see it provide good coverage, as it demonstrates that the team is able to detect events based on the telemetry collected.

It is important to note that the numbers for the delayed detection don't necessarily mean it was the only way a detection was or could be made; the same detection could be identified by other means. Our results show the strength of our MDR analysts, however, our detection coverage results would have remained strong without this human involvement—approximately 86% detection coverage.

Major Step Breakdown



Trend Micro's correlated detections seem low, why?

The priority category for this evaluation is the main detections. Correlation falls into the "modifier detection" category, which looks at what happens above and beyond an initial detection.

Given that our XDR platform was not part of this evaluation, nor did this evaluation assess correlation across security layers–email security, for example, was not in scope for this evaluation–there is correlation value we can deliver to customers beyond what is represented here.

Where did we rely on Deep Discovery Inspector?

Most activities related to this APT29 simulation happened within the host itself, so there were only had a handful of things related to the network.

Unlike many competitors' products, Apex One already sees network events from its vulnerability protection host-based intrusion prevention system (HIPS) features and web reputation kernel hooks. Deep Discovery Inspector and Apex One both detected a number of network activities, and ultimately, Deep Discovery Inspector provided two unique detections. Our results would have remained very strong without it—approximately 90.2% detection coverage.

Why was Deep Security included?

The attack scenario included some Windows servers, so Deep Security was in scope for this evaluation. As well, because Deep Security is on an Active Directory server, it had some visibility into the endpoint activity through log inspection into active directory logs.

What things should be considered when looking at Trend Micro's results? What does the test not consider?

It is important to recognize that this evaluation only tests detection after an event. As a result, it does not measure or account for any automated detection and response measures. This is significant for Trend Micro, as our philosophy is to block and prevent as much as possible, so customers have less to clean up and resolve. Our product was not originally designed for this kind of testing parameter, so in some cases we don't record an activity (have a detection-only option) because we would have blocked it. This test does not show that.

In at least 10 steps of the targeted attack, automated detection and response would have intervened, interrupting the attack with a blocking action (kill process, quarantine, isolate, and etc.).

As well, in the case of many APTs, it would likely have started with a social engineering effort (e.g. phishing email), which would have likely been detected earlier by our email security. MITRE ATT&CK does not test at the point of initial access, so security at this stage of the kill chain is not evaluated.

It is also important to note that this evaluation only looked at endpoints and servers on Windows; it did not look at Linux, where of course Trend Micro has advantages over other vendors.

No other product considerations are included in the evaluation, for example, it doesn't evaluate performance, false positives, cost of ownership, integration with other tools, user interface, security policies, product vison, roadmap, and other factors that can be important selection criteria for organizations.

Is there a winner or leader for the results? Why are other vendors saying they scored best, evaluated highest, etc.?

MITRE ATT&CK does not score, rank, or provide side by side comparison of products. Vendors must take the raw results and parse the data to calculate scores and understand their results in comparison to other vendors. There are very many ways to assess the data, some more valuable and straightforward than others.

We had very good results overall, and scored strongly across many of the parameters tested and according to many of the ways the data can be viewed. With over 91% detection coverage, we showed a great balance of detection capabilities. Our results also support our strategy of providing higher fidelity detections, while managing alert volumes.

How does Trend Micro XDR apply to this MITRE ATT&CK evaluation?

The XDR platform that will be made available in late June 2020, was not part of the evaluation. There are a number of XDR advantages:

- Correlated detections based on rules that look for different behaviors across security layers
- Fewer actionable alerts with greater MITRE ATT&CK context
- Visibility and integrated investigation across security layers—endpoint, email, network, servers, and cloud workloads

OTHER MITRE ATT&CK INITIATIVES WITH TREND MICRO

What else do we do with MITRE ATT&CK?

Trend Micro works closely with the MITRE ATT&CK organization and contributes regularly to the MITRE ATT&CK TTP framework, sharing any new techniques not currently listed their matrices. Both Trend Micro and Trend Micro™ Zero Day Initiative™ (ZDI) are Common Vulnerability Exposures (CVEs®) Numbering Authorities (CNA). Trend Micro can issue CVEs for vulnerabilities specifically for Trend Micro products (they don't have to be found by us). ZDI can issue CVEs for products that are not already covered by another CNA. We are on a number of working groups with MITRE ATT&CK and other peers in the industry working on vulnerability related issues.

What product capabilities does Trend Micro have in relation to MITRE ATT&CK?

- Apex One maps detection logs to MITRE ATT&CK tactics, techniques, and procedures. There are several mappings that can be seen from either the detection dashboard or detection log.
- Trend Micro Cloud One™ and Deep Security reference MITRE ATT&CK IDs directly in the intrusion prevention system (IPS), integrity monitoring, and log inspection rules. In the future, we plan to publish a heatmap of our coverage.
- Trend Micro™ Tipping Point™ references MITRE ATT&CK IDs in rules, and provides extended mapping in the references .xml file (commonly known as dvreferences.xml).
- Deep Discovery also maps detection results to the MITRE ATT&CK Matrix tactics and techniques to provide greater insight and visibility into the methods and vectors used across the attack life cycle.
- Trend Micro™ Cloud App Security maps detections to techniques in the Cloud (Microsoft 365®) MITRE ATT&CK Matrix.
- Trend Micro™ XDR includes:
    - Specific MITRE ATT&CK framework views to map with entire MITRE ATT&CK canvas on all enriched detection and telemetry.
    - Hunting with individual techniques, sequentially described custom criteria (collection of specific attack techniques), and SIGMA file support.

Now that you understand the results and framework, it's important to know how you can leverage the information.

Many organizations are starting to use MITRE ATT&CK to improve their security operations by making sure they are adequately equipped for common adversarial behavior. You can use it to assess your processes and capabilities to identify any gaps, but also overlaps in coverage that may be causing extra operational costs. If you do find those gaps, you can look at evaluations, like this one, to see which solutions can help fill them. As the threat landscape continues to shift and evolve, MITRE ATT&CK provides a common language and context for companies to ask vendors about their capabilities, making it easier to cut through the noise and get the best security.

TREND MICRO™

Securing Your Connected World

[Asset01_Trend_Responds_MITRE_ATTACK_EVALUATION_200514US]