WHITE PAPER

Top Considerations for ML-Based Cloud Security

threat stack
WITH threatML™

Introduction

Many are popularizing the belief that machine learning (ML) is a magical, quick fix for all security use cases. The reality is computers are good at doing math, but humans are needed to properly contextualize their calculations to make them meaningful. Thus, valuable ML learning models don't operate in a vacuum, they rely heavily on relevant data and human input.

Incorporating ML into your cloud security practice requires a strong understanding of the specific problem targets you are aiming to solve and what data sets align to those problems. It is also critical to leverage a skilled team of cloud security experts, data scientists, and engineers that can help correlate problem targets with relevant data sets to derive valuable insights.

It is also important to note that while improved risk and anomaly detection can reduce alert volume, that should not be the primary goal. Instead, look for ML to add valuable context to your security alerts and their associated rules.

A "human in the loop" approach takes into account the unique needs of your IT environment and incorporates your existing security data and alerting rules, which can greatly accelerate your ML-based cloud security strategy. As a result, you can more proactively uncover risk and make better decisions for securing your critical infrastructure and applications.

Top Misconceptions for ML-Based Cloud Security:

01

ML can solve all security use cases.

Repetitive and manual tasks are great for automation, but human expertise is needed to determine appropriate use cases. Gain real value from ML by leveraging a solution built and supported by domain experts and services.

02

You can get by with limited data.

As the infrastructure and application stack becomes more layed, security telemetry becomes foundational for holistically assessing risk and building effective ML-models. Leveraging relevant security information produces more meaningful insights and actionable alerts.

03

ML alone is enough.

Computers know statistics, but humans are needed to pre-define threats that are specific to their environment and interpret ML findings and determine their relevance in order to make ML meaningful.

Top 4 Steps to Get Started

STEP 1 →

COLLECT THE RIGHT DATA

STEP 2 →

ESTABLISH USE CASES

STEP 3 →

ASSEMBLE THE EXPERTS

STEP 4 →

PLAN FOR CHANGE

STEP 1 — STEP 2 — STEP 3 — STEP 4

1 Collect the Right Data

As workloads become more dynamic and complex, security data is becoming increasingly fragmented across cloud management consoles, virtual machine hosts, containers, and applications.

It is essential to collect telemetry from all of these sources, as it provides a wealth of relevant information to create meaningful ML-models.

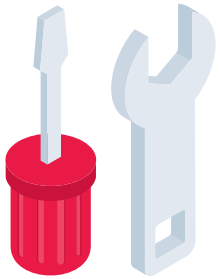
Machine learning models can accelerate anomaly detection, but without investing in broad telemetry collection and rules creation, the alerts created by anomalous activity will lack critical context. Similarly, with ML anomaly detection, security analysts can be made aware of suspicious trends, resulting from a wide range of activities that, in and of themselves, may not trigger an alert. But when these activities are grouped and looked at holistically, they can uncover previously undetected threats.

EFFECTIVE ML-MODELS TAKE INTO CONSIDERATION:

- Who has accessed sensitive data
- Which files have they touched
- Who can log into production systems
- What commands did they run once they logged in

STEP 1 — STEP 2 — STEP 3 — STEP 4

2 Establish Use Cases



While ML is often advertised as a silver bullet for all security issues, determining good candidates for ML is not always as obvious as one may think. Let's consider the case of detecting root escalations. While it is a seemingly simple use case, it can actually be expanded to multiple technical requirements.

For example, an organization may seek to send an alert every time a user executes a command as a user withID of 0. However, this is typically not a well-defined problem and solution. Instead, it's likely more preferable to automatically disconnect a system where root escalation was detected, before that user has a chance to run a destructive command. And they may not want to do this for every non-root user (e.g., some users may require elevated privileges to perform their job), so discovering when a user elevates their own access privileges provides useful context to this scenario.

To get started on your ML-based cloud security journey, it is helpful to identify a core use case to solve for. Recent compliance audits and security reports are typically ideal sources for identifying the most business impactful security issues to solve.

Similarly, not all anomalies indicate risk and require alerting. This is especially the case in cloud and containerized environments with workloads that can be deployed and decommissioned flexibly and have broader accessibility. In these dynamic environments, activities deemed "normal" versus "anomalous" can change frequently.

> "The adoption of AI can have a very positive impact on an organization's security posture and bottom line. The biggest benefit is the increase in speed of analyzing threats."
>
> -Ponemon Institute

Having a team of experts can help you determine if a problem is currently unsolvable, but could be with more quantifiable investment, or if the problem is solvable, but other indicators have to be accounted for in order to gain the desired insights.

STEP 1 — STEP 2 — STEP 3 — STEP 4

3 Assemble the Experts

Even the best security telemetry and rules can't replace the knowledge, reasoning, and decision making skills of humans. The most effective ML models incorporate deep cloud infrastructure and application security expertise, including knowledge from highly skilled security solution architects and SOC analysts.

It is also important to have data scientists and engineers involved throughout the process. They are essential

for extracting context from the more abstract descriptions of the problem the organization wants to solve. Data scientists organize, tag, and catalog data so that it provides meaningful insights related to the investigation and remediation of the problem at hand. While data engineers identify and move the data required to implement a solution such that data scientists, with the help of subject matter expertise, can efficiently analyze it and develop the required algorithms and methodologies.

Once the ML model is in use, context regarding how security teams are interacting with its associated platform in terms of alert dismissals, escalations, and/or rule modifications must be used to enhance the system's anomaly detection and scoring capabilities as well as help inform alerting rules over time.

This human-machine feedback loop results in ML models that effectively drive down the cost of security investigations by improving mean-time-to-detect (MTTD), mean-time-to-know (MTTK), and mean-time-to- respond (MTTR).

> "The shortage of technical security staff, the rapid migration to cloud computing, regulatory compliance requirements and the unrelenting evolution of threats continue to be the most significant ongoing major security challenges."
>
> -Gartner

STEP 1 — STEP 2 — STEP 3 — STEP 4

4 Plan for Change

It is important to recognize that advanced use cases will emerge and organizational requirements will change over time. Therefore, ML investments must take adaptability and reusable components into account.

Take for example detecting anomalous behavior on a host after a few failed attempts at logging into the AWS console. This scenario will require a machine consumable representation of the data stream that captures temporal relationships between various data sources. In this example, that would be CloudTrail and operating system audit events.

> "In response to the dynamic nature of the immediate threat landscape, Gartner recommends that organizations invest in security solutions that are agile enough to evolve alongside it."
>
> -Gartner

Further considerations also need to be made on how "anomalous" behavior is defined. For instance, should the user behavior be considered in the context of day of the week, or a month, or employment tenure of the user? Each context requires additional data sources and investment.

Over time, by working with your team of ML experts, your organization can amass reusable building blocks that will allow it to more efficiently leverage its cloud security investments, to include ML, to solve such problems.

ML (Done Right) is a Team Sport

It is also important to take a strategic and unified approach to allocating your resources. Otherwise, some parts of this journey will become unnecessarily fragmented, time consuming, and costly. For instance, the misutilization of internal staff as subject matter experts and/or the accumulation of multiple point products is what stalls and/or derails most ML projects.

Likewise, even the best rules and ML technology can't replace human intuition, reasoning, and decision making skills.

Threat Stack leverages multiple sources to train our ML models, to include the deep cloud infrastructure and application security expertise of our in-house security solution architects and well-trained SOC analysts. Threat Stack with ThreatML™ encodes proven cloud expertise gained through our Oversight℠ and Insight℠ professional services, which includes the coverage of hundreds of customers and their unique production cloud environments. This experience underpins Threat Stack's ML-based anomaly detection capabilities.

The more our experts work with your team, and you interact with the platform, the more ThreatML models learn over time — adding valuable context that improves future findings.

Learn how Threat Stack's human guided approach to ML can accelerate your cloud security journey →

Successful ML-based cloud security initiatives place a strong emphasis on identifying valuable problem targets and correlating them with relevant data sets to architect meaningful solutions. They also infuse human expertise and decision making skills throughout the process. This is an iterative technique that requires a number of important elements, mostly notably:

- ✓ Determining your problem targets (what are you trying to solve)
- ✓ Identifying what data is required to provide relevant insights
- ✓ Establishing a team of seasoned security professionals, to include security professionals, engineers and data scientists
- ✓ Forming a continuous feedback loop for improving the ML model over time as your environment and policies change

About Us

Threat Stack is the leader in cloud security and compliance for infrastructure and applications, helping companies securely leverage the cloud with proactive risk identification, real-time threat detection, and full stack security observability through the powerful combination of the Threat Stack Cloud Security Platform® and the Threat Stack Cloud SecOps Program℠. For more information or to start a free trial, visit threatstack.com.

55 Summer Street
Boston, MA 02110

threatstack.com | @threatstack