Macula –
ThreatStream's AI Engine

Macula is Anomali ThreatStream's machine learning-based threat-scoring engine. It analyzes, predicts, and classifies domain, IP, and URL observables ingested into ThreatStream through various means—premium feeds, open source feeds, Anomali Labs, and more. It uses a supervised learning algorithm to learn and build the rules/knowledge base (the model) from malicious and benign IOC examples (the training set).

Macula continuously processes, analyzes, and classifies observables. With time and a continuously increasing number of observables submitted for processing, it is constantly learning, adapting, and improving its accuracy to analyze incoming intelligence and reduce false positives.

Macula is behind the scenes of Anomali's ThreatStream repository. When you receive threat intelligence from ThreatStream, you receive curated, validated, and scored threat intelligence that has been processed by Macula. Macula also ensures that the intelligence from ThreatStream to your downstream infrastructures such as SIEMs and firewalls is as reliable and secure as possible, reducing the staff fatigue that can result from excessive alerts.

Macula Model

Macula uses the supervised learning model—the process in which training data is correctly labeled and used to teach and train. Once Macula has learned to recognize and classify data based on the training set, it becomes capable of recognizing and accurately classifying unlabeled data.

During the learning stage, Macula considers the attributes that separate malicious from benign observables and builds a set of rules/knowledge base. Anomali's Data Science team carefully chooses, studies, and constantly revises these attributes to ensure and improve model accuracy and predictability.

As an example, for IP-based observables, attributes such as the following are considered:

- Previous reports of malicious behavior
- ASN
- Geolocation
- Open ports
- Number of domains that resolve to the address
- Entropy of domains that resolve to the address

Similarly, for the Domain- and URL-based observables, attributes such as the following are considered:

- Previous reports of malicious behavior
- Protocol (http/https)
- Entropy of domain/URL
- SSL certificates
- Domain registration date
- Domain popularity and ranking
- Domain DNS activity

Sophisticated machine learning methods are used to study the relationship between the maliciousness of observables and the values of those attributes. The results of this learning form the basis of the machine-learning model.

Macula Components

The Macula system includes the following components:

Whitelist

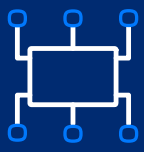
A list of known domains, sites, etc. The whitelist is generated from several well-known sources. The incoming observables are matched against the whitelist; if matched, observables are not scored by Macula. Observables in the whitelist are never added to Anomali's threat intelligence database.

Blacklist

A list of observables known to be malicious. The blacklist is generated from several well-known sources. The incoming observables are matched against the blacklist; if matched, observables are assigned a score of 100 and ingested into ThreatStream.

Telescope

Retrieves and stores raw external data enrichments such as passive DNS information, open ports information, geolocation, online reputations, and more. Macula uses Telescope to extract meaningful information and score observables.



IP and Domain Models

Machine learning models to analyze observables not matched in the Whitelist or Blacklist. These models determine whether observables are malicious at the time of ingestion and Macula's confidence level in doing so. IP address observables are routed to the IP Model; domain and URL observables are routed to the Domain Model.

The models continue to improve as they analyze and classify incoming observables. As a result, Macula's model evolves continuously to become smarter and more accurate as time progresses.

Scoring

Macula gives the observables it processes a Confidence score. This score indicates how confident Macula is that the observable exhibits or is connected to malicious behavior. Confidence scores are from 1-100. A confidence score of 100 indicates that Macula is 100% confident the observable is malicious based on the information available at the time the observable is processed. Lower confidence scores indicate that Macula is less confident in establishing the maliciousness of an observable. Sometimes Macula assigns low confidence scores when observables are new and adequate information is not yet available to assign maliciousness with a high degree of confidence. For observables that belong to your organization in ThreatStream, you can manually edit confidence scores when you know an observable to be malicious. If the observable does not belong to your organization, contact Anomali Support.

Insights Into Decision Making

The Insights feature provides detail on the factors/influencers that were taken into consideration when classifying the observable. These insights offer the transparency customers expect to better understand the rationale behind the classification.

Improving Model Accuracy

Macula continues to learn from its own misclassified and correctly classified information. If its initial claim for an observable is found to be misclassified or incorrectly scored, Macula has the ability to learn and improve its predictive power. The ThreatStream platform allows customers to report back misclassified or incorrectly scored observables. The Anomali Labs team in partnership with the Anomali Data Science team works to verify the claim and revise the model as needed. Similarly, customers report My Recent Attacks information back to ThreatStream, which also allows Macula to ascertain when observables predictions were right. This process of constant feedback and refinement helps ensure that the Macula engine continues to become smarter and more accurate in its predictions.

ANOMALI