TRACEABLE

SOLUTIONS BRIEF

Intelligent Rate Limiting

API Rate Limiting and Throttling for Advanced API Abuse Prevention

Introduction: What is API Abuse?

API Abuse has recently become an important topic among security professionals, and for good reason. In the past two years, we've seen large scale data breaches happen as a result of APIs being abused and misused in some way.

API Abuse occurs when a malicious party uses an API in a way that was not intended by its original design, such as making excessive requests to a server in order to cause a denial of service attack, or using an API to access sensitive information without proper authorization.

It can take many forms, and the specific type of abuse will depend on the functionality of the API, as well as the nature of the attack.

Some common forms of API abuse include:

- Scraping: using automated scripts to extract large amounts of data from an API, which can slow down or crash the server.
- Spamming: the bulk creation of new accounts, as well as programmatic transactions and bidding at auctions.
- GraphQL APIs: server overload can cause a whole host of operational problems. To protect GraphQL APIs, organizations must prevent this by limiting the number of operations such as, queries, mutations, subscriptions, and more.
- Stealing data: Using the API to access sensitive information like personal user data or financial information.

SOLUTION HIGHLIGHTS

- **Robust Rate Limiting:** Implement multiple rate limiting strategies to prevent abuse.
- **Authentication and Authorization**: Require API keys, OAuth, or other secure methods for access control.
- **Anomaly Detection**: Monitor and flag unusual requests from known malicious IP addresses.
- **Source-based Controls**: Different limits can be set for bots, anonymous VPNs, residential proxies, data centers, and more.
- **Geofencing**: Restrict access to the API, based on the client's geographic location.
- **Integration with Security Tools**: Seamless integration with existing security systems such as WAF and SIEM.

Intelligent Rate Limiting for API Abuse Prevention

Malicious API attacks have increased 137% in 2022 alone, and despite the security measures in place, they are predicted to become the primary attack vector in 2023.

We have already seen that to be true with data breaches seen at T-Mobile and an ethical hacker's disclosure of vulnerabilities he was able to exposed at Toyota.

It is evident that InfoSec teams need to tighten API protection to boost overall security posture. API throttling and rate-limiting are two important capabilities for managing APIs that need dedicated protection.

The Increasing Need for Intelligent Rate Limiting

As API traffic volumes increase manifold businesses have to look at multiple means to impose rate limits on their APIs. The need for doing that stems from:

1. Prevent system overwhelm – The goal of any API backend is to provide a high-quality service for all client requests. If a single client floods the server with requests that basic need cannot be met. The sudden or sustained spikes in API traffic will impact that API and all downstream services which that API is front ending in most cloud native environments.

2. Breach Service SLAs – APIs must be able to handle specific amounts of traffic to meet the needs the clients which rely on these APIs have to sustain their business needs. To do this, typically system owners control their client's rates, so they stay within expected bounds for service-level agreements (SLA) which they have agreed with their clients.

3. Operating expenses and cloud costs – Most APIs are hosted in the cloud these days and they can consume large amounts of resources all the way from computer to network to storage costs. This is also true when API's have dependencies on 3rd party APIs which are charged based on consumptions like Salesforce, Twilio or Plaid APIs.

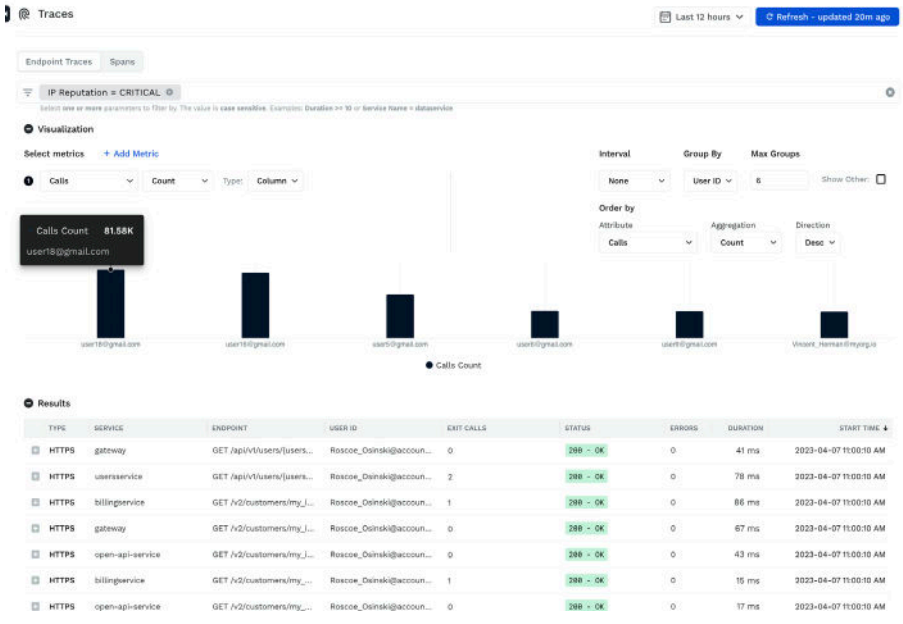
ADDITIONAL FEATURES

- **IP Reputation Scoring:** Block or limit request patterns or usage spikes.
- **Machine Learning:** Use AI algorithms to detect and block potential abuse patterns.
- **Customizable Rules:** Allow configuration of specific rules to suit business requirements.
- **Logging and Monitoring:** Capture detailed usage logs for analysis and real-time monitoring.
- **Reporting and Alerting:** Provides comprehensive reporting and timely alerts for potential abuse incidents.

Identify the Right APIs and Attributes for Rate Limiting

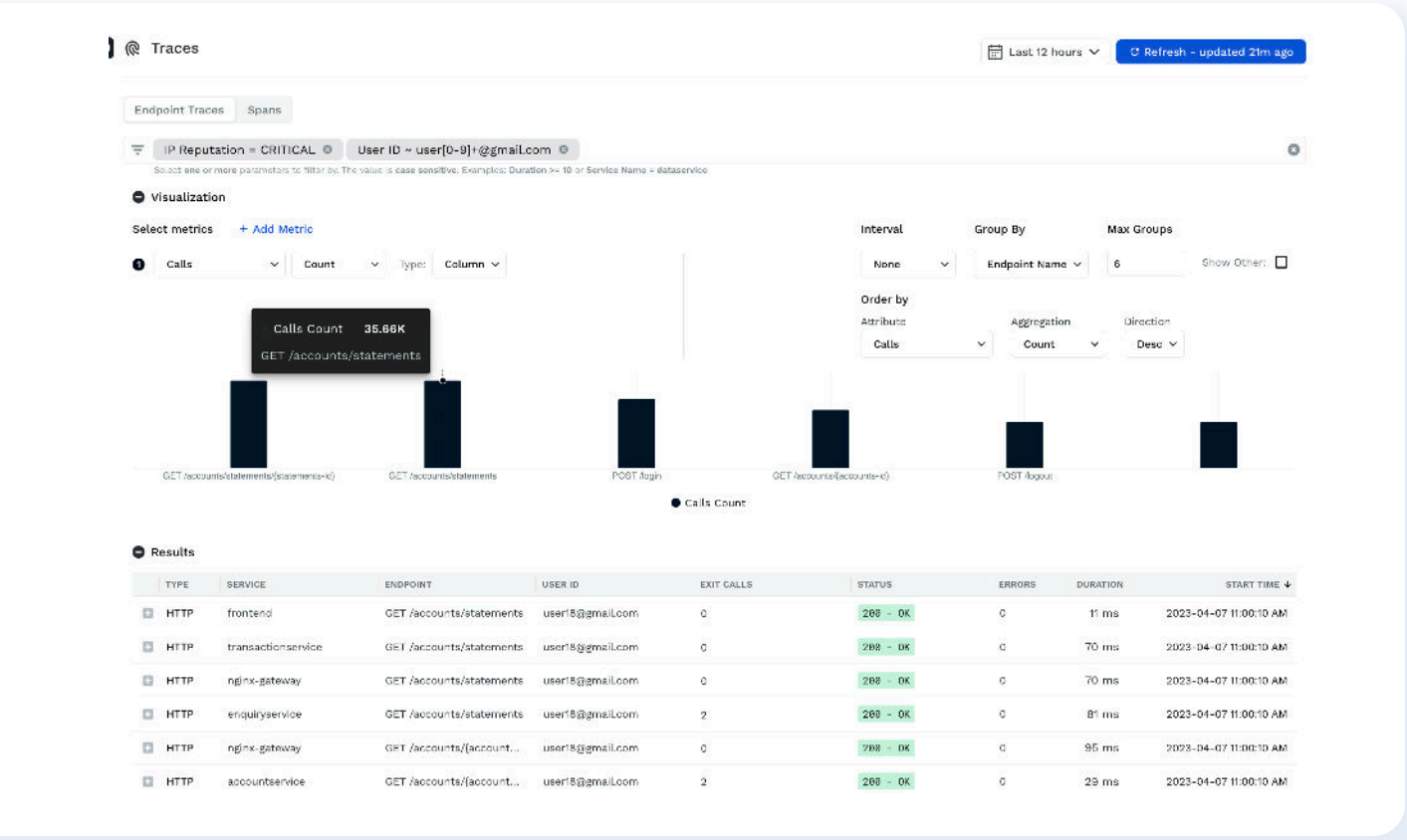
Identify the Right APIs and Attributes for Rate Limiting

"Interesting" APIs – APIs which are critical to the business function and can severely impact continuity when overwhelmed need to be identified upfront. In few cases, the business is aware of such APIs and their capacity – this is the easy but rare scenario. In majority of the cases, businesses struggle to capture the "interest" metric in their APIs. That's where Traceable can help by using the security analytics functionality by providing all the API endpoint traffic and merging that with information of sensitive data, authentication, source of the traffic (BOTs, Residential proxy, TOR, Anonymous VPN's etc).



"Suspicious" Users – Sometimes a group of users target a few endpoints with some malicious intent. Identifying and stopping them is very crucial in the smooth functioning of the business.

Traceable helps in identifying such users or user groups as well as their intended target API's and sensitive data (PCI, PII, HIPAA, GDPR etc) or materially significant data they access (payment info, referral amounts, gift card numbers, etc) that are specific to the business.

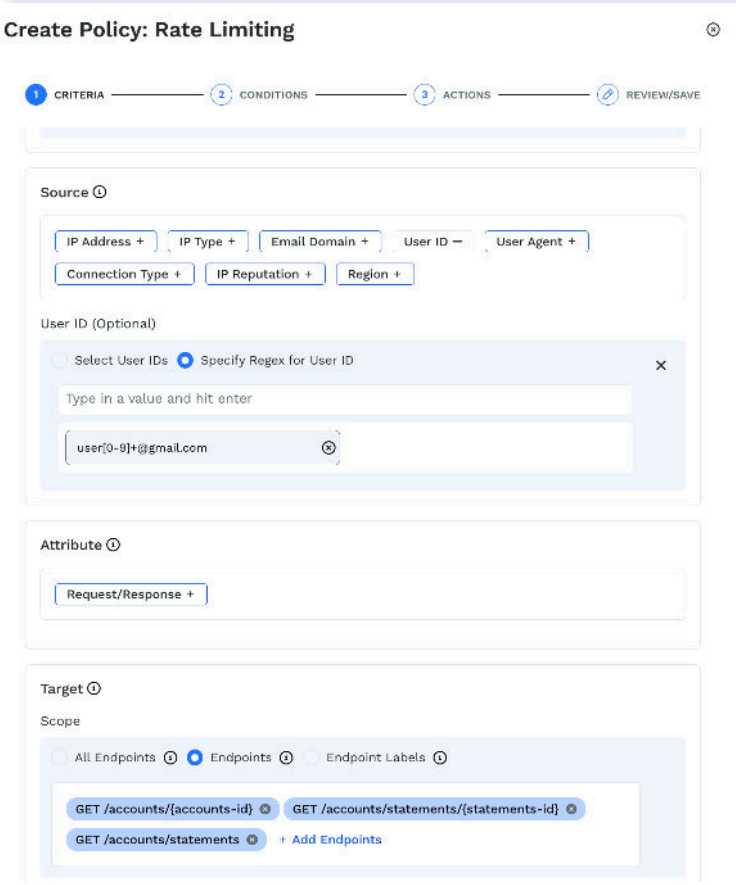


Apply Policies for Rate Limiting

Traceable allows security teams to apply policies to rate limit APIs and/or users by defining known candidates or broader options in their absence, defining crucial conditions to identify the targets where rate limiting is essential for the business applications to function smoothly.

Along the same lines, Traceable allows options to define rate limits manually or with GraphQL APIs which is useful when the intended volume of the endpoints is known. Traceable also provides options to dynamically calculate the rates based on usual traffic and use that to rate-limit endpoints and/or users.

Businesses can apply these policies to detect overuse of endpoints or to even block such abusing users, source types like Bots, Residential proxies, user agents, domains etc from accessing the application.



Custom Rate Limiting Policies

Along the same lines, Traceable allows options to define rate limits manually or with GraphQL APIs which is useful when the intended volume of the endpoints is known. Traceable also provides options to dynamically calculate the rates based on usual traffic and use that to rate-limit endpoints and/or users.

Organizations can apply these policies to detect overuse of endpoints or to even block such abusing users from accessing the application.

The Bottom Line

Attackers have been specifically targeting mobile and Web applications to hijack API calls and to use automated attacks on critical business APIs. They target APIs to take over accounts, steal tokens, scrape business-critical data, and perform application distributed denial of service (DDoS) attacks.

API Rate limiting using contextual information on normal API access rates per source type is the best way to understand when your APIs have been targeted to take the necessary remedial action.

At Traceable we started with the data lake approach for all APIs (external and internal) and all users (malicious, suspicious, benign) to protect you from these sophisticated attacks.