Generate a secure policy for API rate limiting . Ensure security best practices.

API Rate Limiting Policy with Security Best Practices

**1. Purpose**

Prevent API abuse, ensure fair resource allocation, protect against DDoS attacks, and maintain service availability by enforcing controlled access to APIs.

**2. Scope**

* **Applicable To**: All public/internal APIs, third-party integrations, and authenticated/unauthenticated users.
* **Exclusions**: Critical health-check endpoints (e.g., /status).

**3. Rate Limits**

| **Tier** | **Limit (Requests)** | **Time Window** | **Scope** |
| --- | --- | --- | --- |
| Unauthenticated | 100 | 1 minute | Per IP address |
| Standard User | 1,000 | 1 minute | Per API key/user token |
| Premium User | 10,000 | 1 minute | Per API key |
| Critical Endpoints\* | 50 | 1 minute | Per user/IP (whichever is stricter) |

\*Applies to high-risk endpoints (e.g., login, password reset).  
**Note**: Limits auto-adjust during traffic surges (e.g., +25% during peak).

**4. Client Identification**

* **Authenticated Requests**: API keys/OAuth tokens (validated via JWT signatures).
* **Unauthenticated Requests**: IP address + User-Agent fingerprint.
* **Security Controls**:
  + Reject malformed/missing credentials (HTTP 401/403).
  + Use short-lived tokens (max 1-hour expiry).
  + Rotate API keys quarterly.

**5. Exceeding Limits: Response Protocol**

* **HTTP Status**: 429 Too Many Requests.
* **Response Headers**:

http

Copy

Download

Retry-After: 60 # Seconds

X-RateLimit-Limit: 1000

X-RateLimit-Remaining: 0

X-RateLimit-Reset: 2025-05-27T10:00:00Z

* **Security Measures**:
  + **Progressive Penalties**:
    - 1st violation: 1-minute cooldown.
    - 3+ violations: 24-hour block + security alert.
  + **No Sensitive Details**: Error messages omit system internals (e.g., "Rate limit exceeded").

**6. Monitoring & Logging**

* **Log All Requests**:
  + Timestamp, client ID, endpoint, response code.
  + Anonymize IPs (store only /24 subnet).
* **Real-Time Alerts**:
  + Trigger for:
    - 5 clients blocked in 5 minutes.
    - Repeated 429 from a single user.
* **Retention**: Logs retained for 30 days (GDPR-compliant).

**7. Exceptions & Escalation**

* **Whitelisting**:
  + Internal microservices (authenticated via mTLS).
  + Approved partners (contractually bound).
* **Limit Increases**:
  + Submit request via **IT Security Team** with:
    - Business justification.
    - Threat model assessment.
  + Approval requires CISO/CTO sign-off.

**8. Security Best Practices**

* **Defense-in-Depth**:
  + Layer 1: Edge firewall (block known malicious IPs).
  + Layer 2: API gateway rate limiting.
  + Layer 3: Service-level quotas.
* **Token Revocation**: Immediate invalidation for leaked/compromised keys.
* **Encryption**: All rate-limit counters in transit/in rest (AES-256).
* **Bot Mitigation**: CAPTCHA challenge after 3 failed auth attempts.

**9. Policy Review**

* **Annual Review**: Adjust limits based on usage telemetry and threat landscape.
* **Trigger-Based Review**: After security incidents or infrastructure changes.
* **Compliance**: Align with OWASP API Top 10, NIST SP 800-204.

**Approved by**

* Chief Information Security Officer (CISO)
* API Engineering Lead
* Legal & Compliance Office

**Effective Date**: 2025-06-01  
**Version**: 1.2

🔐 **Implementation Tip**: Use tools like *AWS WAF*, *Kong API Gateway*, or *Redis* with sliding-window counters for scalable enforcement. Test limits via chaos engineering (e.g., simulated DDoS).