

Maskbreak Security Whitepaper

Sentinel Edge Networks LTD · Registered in England & Wales 17150600 · 134a West Hendon Broadway, London NW9 7AA · Version 2.5 · 18 October 2026 · canonical copy at maskbreak.com/security-whitepaper

Maskbreak is a real-time fraud detection API (maskbreak.com). This document states our security posture plainly for procurement and security review: the controls in place today, how data is handled, and — deliberately — what we do not yet have. It is written to be checked, not to impress: every claim here is verifiable on the linked public pages, and we do not list controls we don't have.

1. Certifications & audits — the honest version

Maskbreak is **pre-audit for SOC 2 Type II** and holds no SOC 2, ISO 27001, HIPAA, or PCI DSS attestation today. No third-party penetration test has been commissioned yet. We say this here rather than imply otherwise with badge walls; this section will be updated as each milestone lands, with the change dated on our changelog. A responsible-disclosure program is active (</responsible-disclosure>, RFC 9116 [security.txt](#) published).

2. Technical controls in place today

CONTROL	IMPLEMENTATION
Authentication	Email + bcrypt-hashed password (new passwords longer than bcrypt's 72-byte limit are refused rather than silently truncated), Google OAuth (RS256, JWKS verified), GitHub OAuth, TOTP 2FA with recovery codes, passkeys (WebAuthn), reviewable active-session list with per-device revocation. Sign-in is locked after 10 failed passwords in 30 minutes per account, and password re-checks before sensitive changes count toward the same lock; wrong two-factor codes have a per-account daily limit, and the account holder is emailed when a correct password is followed by wrong two-factor codes. Only genuine console-session tokens authenticate the console; other signed tokens we issue (OAuth state, admin and API access tokens) are refused there. Since 14 October 2026 the console session is held in an HttpOnly, Secure, SameSite=Strict cookie, which page scripts cannot read; the server accepts it only on requests carrying the console's own header, so other sites cannot use it. A browser that refuses cookies keeps the session token in page storage, as before.
API key management	Keys generated with crypto.randomBytes (prefix <code>sk_live_</code>); rotation and instant revocation via dashboard; key rotation and password change require re-authentication.
Password breach check	Have I Been Pwned k-anonymity (5-character SHA-1 prefix only) on signup and password reset — the password itself never leaves our infrastructure.

CONTROL	IMPLEMENTATION
Transport security	TLS 1.2+; HSTS for one year with includeSubDomains and the preload directive (the domain has not yet been submitted to browsers' preload lists); CORP same-origin; CSP with allowlists and live violation reporting on every page, including HTML served under /api/ ; a strict first-party policy without inline script is enforced on the internal admin console, and a hash-based policy without ' unsafe-inline ' has been enforced on the customer console since 6 October 2026; X-Content-Type-Options nosniff; X-Frame-Options DENY; Permissions-Policy locks camera/mic/geolocation/payment.
Rate limiting	Per-account monthly allowances and hourly caps set by the plan (the live key and its predecessor share them during a key rotation; the test key has its own hourly cap) and per-source-IP caps on the evaluate and lookup endpoints; fixed-window rate limits on authentication and form endpoints, with IPv6 addresses also counted per /64; per-account limits on signed-in console and API routes; a 30-minute account lockout after 10 failed password attempts; in-memory limiter state is bounded; every 429 response carries Retry-After . The API-key IP allowlist, when configured, also applies to minting OAuth tokens (/oauth/token); /v1/usage stays readable from outside the list, as documented.
Admin access	The internal admin console signs in only through a single-sign-on provider with a phishing-resistant passkey (since 14 October 2026; Google sign-in and the admin key are switched off and come back only through a documented break-glass setting), sits behind an access gateway that admits only the owner on an enrolled device, runs under an enforced Content-Security-Policy without inline script, and writes an audit line to the application log for every admin API request (operator type and email, route, target account id, status and operator IP). Key rotations, two-factor removal, forced sign-outs and reinstatements made by an operator are emailed to the customer; admin sign-ins, key sign-ins and lockouts alert the owner.
Runtime	The server runs in a container as a non-root user on a pinned Node.js version. It refuses to start with known-weak secrets in production, and refuses to connect to the production database from outside the production hosting environment unless explicitly overridden. Outbound calls to providers have timeouts.
Session invalidation	Password change/reset or admin suspension immediately invalidates all sessions (token-epoch bump).
Encryption at rest	Managed encryption at rest on the primary database (Turso / libSQL, AWS us-east-1) and on the backups in Cloudflare R2 (EU jurisdiction, since 17 October 2026; the copies made up to 16 October 2026 in AWS S3, Stockholm, used S3 server-side encryption). Account IDs a customer sends are also encrypted in the database (AES-256-GCM, a key per customer derived from a key held outside the database), since 18 October 2026.

CONTROL	IMPLEMENTATION
Error monitoring	Since 24 September 2026 the server reports unexpected errors to Sentry (EU region, Frankfurt, Germany) through @sentry/node: errors only — no performance tracing, profiling, log forwarding, metrics or session replay, and no Sentry code in the browser. Default PII collection, local variables and trace propagation are off. Before sending, the server drops request headers, cookies, bodies, query strings, the user context and console breadcrumbs; replaces IP addresses, email addresses, UUIDs, purely numeric ids (account and record numbers), hexadecimal strings of 12+ characters and long token-like path segments with placeholders (<code>:ip</code>, <code>:email</code>, <code>:id</code>, <code>:token</code>, e.g. <code>/v1/lookup/:ip</code>); keeps only the method, status code and origin (scheme, host, port — no path or query) for up to 30 outgoing-call breadcrumbs, so a customer's webhook host name can appear; and scrubs API keys, webhook secrets, JWTs, bearer tokens, email addresses, IP addresses, database credentials and the exact values of our own secrets from error messages and the remaining event context (stack frames, which hold our own code, are sent as they are). This scrubbing is pattern-based: a strong safeguard, not a guarantee. In Sentry, IP-address storage is turned off, the server-side data scrubber is on as a second layer, and generative AI features (including Seer) are off. Without a configured DSN the SDK is not loaded at all. Retention follows Sentry's plan: 30 days on the free Developer plan; up to 90 days on paid plans and during a trial.
Uptime transparency	Self-hosted status page at <code>/status</code> with recorded probe samples (90 days) and incident history, plus independent scheduled checks from GitHub Actions (scheduled every 15 minutes; GitHub runs it every few hours in practice) and a Cloudflare schedule (every five minutes). Recorded samples are not a complete uptime measurement; see the service level objectives.

3. Data handling & retention

DATA	POLICY
Lookup IPs	Raw IPs retained 7 days, then replaced by a keyed one-way hash (per customer, key held outside the database) by a cleanup that runs every six hours. Records hashed before 28 September 2026 used an earlier, unkeyed hash; the same cleanup re-keys them with the secret key, and backup copies made before that keep the older form until they expire (60 days).
Application logs	Kept by our hosting provider (Railway, United States) for up to 30 days. Lines identify accounts and records by number and API keys and tokens are not logged on purpose; some lines can still contain an email address or IP address (admin lines name the operator, a failed newsletter send names the recipient, and a request path or error message can contain one).

DATA	POLICY
Backups	A nightly GitHub Actions job (GitHub-hosted runner) reads the full database, compresses it and uploads it to Cloudflare R2 in Cloudflare's EU jurisdiction, encrypted at rest, where a rule removes each copy after 40 days. The job holds no storage key: it writes through a Cloudflare Worker that admits only a GitHub-issued identity token for this repository's main branch, accepts only new copies and deletes nothing. Until 16 October 2026 the copies went to AWS S3 in Stockholm (eu-north-1), encrypted with server-side encryption; that account is suspended, so we can no longer check or delete those copies ourselves. See the retention-assurance note below.
Signup emails (API signal)	Checked transiently against disposable-domain feeds; never stored or logged.
Device and account linking	Per customer only, never linked across customers. Device links use one-way hashes. Since 18 October 2026 an accountId a customer sends is kept as a keyed hash for linking and, encrypted with that customer's own key, only so its console can show it; an account's network addresses are kept as keyed one-way hashes and never for shared exits (VPN, proxy, datacenter, Tor, Apple's iCloud Private Relay, mobile carriers' shared addresses). Deleted 90 days after the account was last checked, and with the customer's account.
Email and phone check (console)	Since 18 October 2026: the one email address or phone number an account holder types is sent to an email- and phone-intelligence provider (United States), with a daily limit of checks per account. Neither the value nor the answer is stored; only the day's count of checks per account (kept 3 days).
Account deletion	GDPR Art. 17 self-service deletion with password re-auth; live-system deletion, including a newsletter subscription under the same address, invitations addressed to the account and pending reset or signup records; the backup-expiry commitment is 60 days; the retention-assurance note below says what the daily audit confirms and what stays open.

DATA	POLICY
Analytics & advertising	Fonts are self-hosted. Google Ads remains removed. No chat widget loads on any page: the Widgo chat was removed on 30 September 2026, and Widgo and the earlier provider Crisp keep legacy records only. Since 23 September 2026 PostHog product analytics records page views, clicks and sessions on public pages, /signup and the dashboard without a consent prompt (skipped under Global Privacy Control, Do Not Track and Save-Data; never on sign-in, reset, verification, invitation, settings, security, members, admin or result pages). Its library (posthog-js 1.434.11, byte-identical to the npm release) is served from maskbreak.com and the CSP grants PostHog’s EU hosts connect-src only, never script-src, so PostHog cannot run code on our pages. Since 17 October 2026 it also receives page-speed measurements (Core Web Vitals) and, on public pages and /signup, clicks that changed nothing and the page’s own script errors, with URLs, IP addresses, email addresses and key-shaped strings removed from their messages. Form fields, the visitor’s own network readouts and all dashboard text are masked in the browser; console logs, network requests and canvas are not recorded; see Cookie Policy §2.3-2.5.

Retention assurance — review dated 17 September 2026, updated 18 October 2026

The existing 60-day backup-retention commitment remains unchanged. The repository now audits all listed current and older backup versions and the storage expiry configuration, without a newest-snapshot exemption. A denied or incomplete inspection cannot pass. Scheduled uploads continue, but a successful upload alone does not establish retention compliance. This audit does not itself erase old versions or change the storage policy. On 10 October 2026 it passed for the first time: an enabled S3 lifecycle rule makes every backup version eligible for deletion within 56 days of its upload, and no stored backup version was 60 days old or older. The audit runs daily and reads what is stored on the day it runs; S3 deletes asynchronously, so it is evidence, not a guarantee of future deletion, and a failed run reopens this item.

Since 17 October 2026 the backups go to Cloudflare R2 (EU jurisdiction), which keeps no older versions: a rule removes each copy after 40 days. The daily audit lists every stored copy through the backup receiver (names, sizes and dates, never contents) and fails if any copy is older than 43 days, so a missing or broken rule shows within days; it first passed on 17 October 2026. The copies made up to 16 October 2026 remain in AWS S3 (Stockholm), where the audit had confirmed the deletion rule from 10 October 2026; that account was suspended between the backups of 16 and 17 October, we are not restoring access to it, and we can no longer check or delete those copies ourselves.

The repository restore tool now blocks downloads and restore instructions because an authoritative record of deletions after each snapshot has not yet been established. It permits a metadata-only inventory. Reopening recovery requires a reviewed deletion-reconciliation process and an isolated restore drill showing that erased accounts and records remain absent. This is a safeguard, not a completed recovery solution. These assurance items remain open, and do not authorise longer retention or restoration of deleted accounts.

4. Sub-processors

Cloudflare (CDN/edge; since 17 October 2026 also the encrypted nightly DB backups in R2, EU jurisdiction, each copy removed after 40 days within the 60-day commitment a daily audit checks), Railway (hosting, United States, US East), Turso (database, United States, AWS us-east-1, including keyed abuse-prevention counters), Resend (email delivery, United States), Stripe (payments and subscriptions for paid plans on its own hosted pages; account holders' billing data only, never customer end-user data), PostHog (product analytics and session recordings, EU Cloud, Frankfurt), Sentry (server error reports, EU region, Frankfurt; designed not to receive personal data, through pattern-based scrubbing that is a strong safeguard rather than a guarantee), Widgo and Crisp (former chat providers; legacy records only), Amazon Web Services / S3 (the encrypted nightly DB backups made up to 16 October 2026, EU Stockholm eu-north-1; the account is suspended, so we can no longer check or delete them), GitHub (GitHub Actions for monitoring and the nightly backup job, which reads the full database; and GitHub Sign-In), Google (optional Sign-In, including One Tap on the log-in page; Google Workspace for the support mailbox), flagcdn.com (flag images in the console's test panel), Have I Been Pwned (k-anonymity check), plus network- and device-intelligence providers (United States) and, for the console's email and phone check since 18 October 2026, an email- and phone-intelligence provider (United States), named to customers under DPA. The live list with regions and a dated change log is public at /sub-processors; new customer-data processing sub-processors are announced in that change log at least 30 days in advance, with right to object; the email- and phone-intelligence provider was the exception, switched on the day of its entry (18 October 2026), and the 30-day right to object still applies to it.

5. Compliance framework

UK GDPR / EU GDPR aligned (processor role for evaluation traffic; DPA with SCCs published self-serve at /dpa for every customer — countersigned copies via support@maskbreak.com). Primary processing is in the United States; EU-only processing is not offered. Section 2 of this document is incorporated into Annex B of the DPA by reference. CCPA honored. Security questionnaires (SIG-Lite, CAIQ, bespoke) completed within 5 business days of request. MSA available where click-through Terms can't be accepted.

This document is a plain-language summary, not a contract, except that section 2 is incorporated into Annex B of the DPA; contractual commitments live in the Terms, DPA, and (for enterprise) MSA/SLA. Verify anything here against the live pages: /trust, /privacy, /status, /sub-processors. Questions: support@maskbreak.com.