



SM CRAWLER

TECHNICAL ASSURANCE · BUILD 194

Technical Security & Data-Protection Review

SM Crawler v194 · Local-first architecture, authentication, storage, encryption boundaries, connector security, translation privacy and recovery

SM Crawler keeps its working message history in the customer-controlled installation. It does not copy that content into a central SM Crawler customer-content database.

IMPLEMENTATION-BACKED · CUSTOMER-FACING · 20 AUGUST 2026

This report is an internal technical review of the inspected v194 implementation. It is not an independent penetration test, security certification or legal opinion.

KEY BOUNDARY

Local files use restricted operating-system permissions. v194 does not yet provide application-level encryption at rest or end-to-end encryption.

How to read this review

This document separates implemented controls from deployment assumptions and customer responsibilities. The strongest accurate message is local customer control—not an absolute claim that no message is stored.

IMPLEMENTED	DISCLOSED BOUNDARY	CUSTOMER CONTROL
Local-first workspace; script password hashes; protected sessions; origin checks; isolated workspaces; signed supported webhooks; verified pre-update backup.	No application-level at-rest encryption, no E2E claim, host-dependent remote access, no independent always-on crawler, no independent certification.	Authorized users, OS administrators, connected platforms, permissions, translation consent, host security, disk encryption, backups and retention.

Contents

1	Executive security position
2	Architecture and trust boundary
3	What is and is not stored
4	Encryption and cryptographic controls
5	Authentication and session security
6	Workspace and tenant isolation
7	Connector and credential boundaries
8	Translation privacy
9	Remote access and network exposure
10	Data integrity, updates and recovery
11	Retention, deletion and customer control
12	Threat review and residual risk
13	Customer security checklist
14	Verification evidence
15	Approved client-facing assurance
16	Document control and release boundary

SECTION 1

1. Executive security position

SM Crawler v194 is a local-first system. Its working engine, message records, connector state, preferences and history run from storage controlled by the customer. In the default local arrangement, SM Crawler's operator does not host a central repository containing customers' personal messages.

The word local does not mean that nothing is stored. It means that the working copies are retained on the customer's computer, in that customer's installation and backups, rather than being copied into a central SM Crawler customer-content database. The customer decides who may use the computer and which SM Crawler accounts may enter each workspace.

The boundary is explicit:

- Personal messages and history may be stored locally so the product can display history, deduplicate events, preserve Done state and recover after restart.
- Platform passwords normally used by a person to enter Facebook, Instagram, Microsoft or another service are not requested or known by the SM Crawler operator.
- Revocable platform tokens, app-specific passwords, API keys or connection references may be stored locally when a connector requires them.
- Local data files are protected by restrictive operating-system permissions, but the current v194 files are not encrypted at rest.
- SM Crawler does not claim end-to-end encryption. Platform services, the customer's devices and any customer-enabled processors remain separate trust domains.
- Secure remote access uses HTTPS when configured. Direct loopback use on the same computer may use HTTP because the traffic does not leave that computer.
- The independent always-on private-cloud crawler described as a future direction is not implemented in v194. The current engine depends on the customer's host computer being on and running.

SECTION 2

2. Architecture and trust boundary

```
Customer's authorised platforms
| Official API / webhook / local authorised reader
| Provider TLS where the provider offers HTTPS
v
+-----+
| CUSTOMER-CONTROLLED INSTALLATION |
|                                     |
| Authenticated gateway -> isolated customer workspace |
| Connector runtime      -> local message/history files |
| Browser/PWA           -> board, replies, notes and agenda |
| Optional translation   -> explicit consent + configured API |
|                                     |
| Files: OS permissions 0700 directories / 0600 private files |
| Current limitation: no application-level encryption at rest |
+-----+
|                                     |
| +-- Optional HTTPS remote route while host is running      |
| +-- Customer-created, verified pre-update backups          |
| +-- Optional external translator only when enabled         |
|                                     |
No central SM Crawler customer-message database is in this path.
```

Trust domains

Trust domain	What it controls	What SM Crawler v194 does
Customer computer and OS	Local users, disk, backups, malware protection and administrator access	Stores the working installation and applies restrictive file permissions
SM Crawler workspace	Customer-authorized users, message state, tokens, preferences and derived records	Authenticates remote users and separates workspaces
Connected platform	Original messages, platform account, permissions and retention	Uses only the configured, supported route and does not replace platform retention
Optional tunnel provider	Encrypted remote transport and public routing while the host runs	Supplies an HTTPS route to the authenticated gateway; it is not an independent crawler
Optional translation provider	Text submitted for translation	Receives message text only after owner consent and provider configuration; sensitive-looking text is withheld
Customer backups	Recoverable snapshots, including any local customer content inside them	Installer creates and verifies a dated pre-update archive

SECTION 3

3. What is and is not stored

The local engine requires durable state to provide history and reliable operation. Depending on the connectors and features selected, the customer-controlled data directory can contain message text or message summaries, sender identity, timestamps, links or media references, reply and Done state, priorities, notes, connection records, tokens, translation results, notification subscriptions, diagnostic state and audit records.

Data class	Stored by SM Crawler?	Location and purpose	Important boundary
Message and conversation records	Yes, locally	Customer workspace history, restart recovery, deduplication and board display	Not copied into a central SM Crawler customer-content database
Contacts and identity mappings	Yes, locally when observed or configured	Customer workspace	Original platform may retain separate identity data
Done, priority, reminder and reply state	Yes, locally	Customer workspace	Customer-created working state
Personal notes and agenda information	Yes, locally when used	Customer workspace	Private files use restrictive OS permissions
Platform personal passwords	No	Customer signs in with the platform directly	Operator does not ask for or know the normal personal password
Tokens, app-specific passwords and API secrets	Sometimes, locally	Private workspace environment or token records when required for a connector	These are secrets; they are not the person's normal platform password and are not encrypted at rest in v194
SM Crawler passwords	No readable password	Only a random salt and script-derived password hash are retained	Hashing is one-way verification, not reversible encryption
Sessions	Yes, as protected session state	Signed account-session cookie or random server-checked local access value	HttpOnly and SameSite=Strict; Secure flag when HTTPS is active
Derived translations	Yes, locally when translation is enabled and used	Private translations.json cache	Can be deleted with the item; original remains authoritative
Diagnostic logs	Yes, locally	Engine and gateway diagnostics	Logs can include event metadata and short message excerpts; treat them as customer data
Backups	Yes, when an update is performed or the customer creates one	Customer-controlled backup location	A backup can contain the same local customer data and must be protected accordingly

Who can access local content

Within the intended deployment, content is accessible to authorized SM Crawler users of the relevant workspace and to operating-system administrators who control the customer computer or its backups. Any person or process that compromises that computer, its administrator account or a copied backup may also be able to read the current unencrypted files. Disk encryption supplied by the operating system, strong device credentials and controlled backups are therefore important customer safeguards.

The SM Crawler operator has no designed central customer-message store to browse. This does not remove the independent access or retention of the original platform, the customer's IT administrators, customer-authorized users, backup operators or any optional processor the customer chooses to enable.

SECTION 4

4. Encryption and cryptographic controls

Layer	v194 protection	What it means	What it does not mean
Remote transport	HTTPS when a secure remote route is configured; HSTS is returned on HTTPS requests	Browser-to-gateway traffic is encrypted in transit on that route	It does not encrypt data already stored on disk
Platform transport	Supported connector APIs use provider HTTPS endpoints	Data is encrypted in transit to supported providers	The connected provider remains a separate data controller/processor
Local loopback	HTTP may be used on 127.0.0.1 / localhost	Traffic stays on the same host networking stack	It is not an HTTPS claim for every local request
Password storage	Salted script digest, 64-byte derived result	The password is verified without storing readable password text	Hashing is not encryption and cannot protect a weak password from all offline guessing
Account session integrity	HMAC-SHA-256 signed payload with expiry and session version	Modified session payloads are rejected; password reset/disable can invalidate sessions	It does not conceal local files
Local access session	Cryptographically random server-checked value	Unpaired remote clients do not receive workspace access	Direct loopback requests trust the local computer boundary
Webhook authenticity	Provider-specific HMAC/signature or secret checks for supported inbound routes	Unsigned or incorrectly signed supported webhooks are rejected	It does not prove every third-party connector is available or approved
Tunnel token handling	Private file with mode 0600 ; token is passed through --token-file , not the service argument list	Reduces accidental exposure in process arguments	The token file itself is not encrypted at rest
Customer files at rest	Directories hardened to 0700 ; ordinary private files hardened to 0600	Other ordinary OS users should not be able to read the files	v194 has no application-level at-rest encryption or E2E encryption

SECTION 5

5. Authentication and session security

Password handling

SM Crawler generates a random salt and derives a 64-byte scrypt hash from the submitted password. Only the salt and derived hash are persisted. Timing-safe comparisons are used for stored digests and session signatures. Client accounts initialized with the temporary password **1234** are forced to replace it at first sign-in, and a password reset increments the session version so existing account sessions can be invalidated.

Session controls

- Account sessions carry a user ID, session version and expiry and are signed with HMAC-SHA-256.
- Relevant cookies are **HttpOnly** and **SameSite=Strict**.
- Cookies receive the **Secure** attribute when the request is carried over HTTPS.
- Account sessions default to a 30-day maximum age; administrators can disable a client, invalidating the active session through the stored session version and active state.
- Repeated failed sign-in attempts are throttled after five failures for the observed address.
- Sensitive administration routes are confined to the admin host and owner role.
- Initial owner credentials can be created only from the main computer's direct loopback context.

Local pairing

Phone/tablet pairing uses a cryptographically random, single-use code with a ten-minute lifetime. Consuming it creates the protected local-access cookie, removes the code from the URL and prevents reuse. The durable access value is not placed in the QR link. Remote setup, credential, tunnel and other sensitive control routes remain restricted to the main computer.

Browser and request controls

The gateway and engine remove the Express technology disclosure and send defensive headers including **X-Content-Type-Options: nosniff**, **X-Frame-Options: DENY** and **Referrer-Policy: no-referrer**. The engine also sets a restrictive framing/base/object Content Security Policy and a Permissions Policy. State-changing browser requests must come from the same origin, except narrowly scoped extension routes for the supported Facebook/Instagram add-ons. Cross-site writes are rejected.

SECTION 6

6. Workspace and tenant isolation

Each gateway user maps to a specific tenant record and workspace. A tenant has its own directory, `.env` connection file and child engine. Tenant directories are created with mode `0700`; private descendant files are hardened to mode `0600`. Child engines bind to `127.0.0.1` rather than a public interface.

The gateway selects the tenant from the authenticated session; the browser does not choose an arbitrary tenant directory. Public tenant URLs expose only the allowed webhook and OAuth callback paths. Workspace data APIs are not made available by the public tenant identifier. Proxy headers that could falsely imply an external identity or origin are removed before requests enter the isolated child engine.

Automated isolation tests create two clients with different secrets and messages, then verify that neither HTTP nor WebSocket delivery crosses between them. Disabling an account invalidates its current session without deleting the tenant's preserved folder.

SECTION 7

7. Connector and credential boundaries

SM Crawler's connector catalogue separates declared capability from verified visibility. A platform is displayed as working only when its required credential group, account or deliberately configured lawful view-only bridge exists. Restricted bridges do not gain reply capability and are limited to approved official destinations.

Credential principles

- The customer signs into each platform through that platform's approved mechanism where supported.
- Configuration-status APIs return whether a secret exists, not the stored secret value.
- The static-file server uses an allowlist so `server/.env`, `server/data` and other private files cannot be downloaded as public assets.
- Disconnecting removes the applicable connection credentials or token records but deliberately preserves archived history unless the customer separately deletes it.
- The customer can revoke a token at the platform as well as disconnect it locally.
- Platform eligibility and permission limits still apply. A configured value does not manufacture API permission the platform has not granted.

Inbound authenticity

Meta/WhatsApp webhooks are validated against the exact raw request bytes using the configured app secret. LINE uses its signed request header. Telegram can require its configured webhook secret. Twilio voice webhooks use Twilio request validation. View-only bridge requests require their configured bearer token. Invalid signatures or tokens are rejected before the event is accepted.

SECTION 8

8. Translation privacy

Message translation is disabled unless the owner records explicit consent and configures a supported provider. If consent is absent, the runtime does not call the external provider and does not create a translation cache. Text that appears to contain a password, verification code, API key, private key, secret, payment-card number or similar credential is withheld from translation.

When translation is enabled, the selected message text and language parameters are sent to the configured provider over its HTTPS API. The provider is therefore an additional processor selected by the customer. The translated result, provider name, language metadata, timestamp and a source-text hash are cached locally with mode **0600**. The original text remains authoritative and visible. The derived translation can be deleted with the item. A message already in the reader's selected language is not submitted and does not display a redundant translation status line.

SECTION 9

9. Remote access and network exposure

The local engine trusts a true direct loopback request because it originates on the customer's own computer without forwarded-IP headers. Remote browsers must authenticate or pair before reading the feed, configuration or control APIs. A host-header rebinding request does not gain loopback trust merely because it reaches **127.0.0.1**.

When a named tunnel is configured, its token is copied into a private control file and passed to the tunnel process through a token-file argument, keeping the token out of ordinary process arguments. The tunnel connects the public HTTPS address to the authenticated gateway while the customer's host is running.

This tunnel is a transport route, not an independent private-cloud crawler. If the customer computer, engine, network or tunnel is off, the current v194 system does not continue collecting on a separate SM Crawler server. No 24/7 availability or SLA should be inferred.

SECTION 10

10. Data integrity, updates and recovery

Local JSON writes use a temporary file followed by an atomic rename and then reapply private permissions. If a stored JSON file is malformed or has the wrong shape, the engine sets the unreadable file aside rather than silently overwriting it, then starts with a safe default. History reconstruction deduplicates records and preserves Done state across restarts.

Before an update replaces the active installation, the installer creates a dated backup outside the installation and verifies that the archive can be read. If backup creation or verification fails, the update stops. The installer payload is rebuilt from an explicit allowlist that excludes private `.env` files, customer data, `node_modules` and other non-release state. The builder extracts every generated installer payload and compares its SHA-256 file hashes to the source payload before marking the installer verified.

A recovery archive can contain messages, tokens, logs and all other local state from the protected installation. The customer must therefore protect, retain and dispose of backups with the same care as the live data. The verified backup step reduces update risk; it is not a substitute for a tested organization-wide backup, disaster-recovery and retention policy.

SECTION 11

11. Retention, deletion and customer control

Retention is primarily customer-controlled because the data lives in the customer installation. Marking an item Done changes working state and does not delete the historical record. Disconnecting a platform stops or hides the connection and removes applicable local credentials, but preserved archive data and backups may remain.

For complete local removal, the customer must identify and delete the relevant live workspace files, diagnostic logs, derived translation cache and every retained backup according to its policy. Revoking a token at the original platform prevents further use but does not delete the original platform's copy. Platform-side deletion must be requested or performed under that platform's own controls.

Before destructive deletion, a customer should decide whether legal hold, regulatory retention, contractual evidence or recovery obligations apply. v194 does not claim a centrally operated deletion service because there is no central SM Crawler customer-message database to purge.

SECTION 12

12. Threat review and residual risk

Threat or failure	Implemented mitigation	Residual risk / customer action
Stolen readable password database	Passwords are not stored; salted script hashes are retained	Use long, unique passwords; an offline attacker can still guess weak passwords
Cross-site request from a malicious web page	Same-origin checks reject browser mutations; SameSite=Strict cookies	Protect the local OS/browser and review extension permissions
Unpaired remote viewer	Server-checked session required; one-use pairing; protected cookies	Revoke/disable accounts and protect paired devices
Forged webhook	Provider signature/secret validation on supported routes	Protect webhook secrets and rotate them if exposed
Cross-tenant access	Session-selected tenant, separate directories and loopback child engines	OS administrators still control the host; keep the host hardened
Token exposure in UI	Status routes return presence, not value; public files are allowlisted	Tokens remain readable to a compromised OS administrator because at-rest encryption is absent
Interrupted write	Temporary-file/rename pattern; malformed files are set aside	Hardware/filesystem failure still requires a protected backup
Unsafe update	Verified dated pre-update backup and payload SHA-256 comparison	Test recovery and retain backups according to policy
Sensitive text sent for translation	Explicit owner consent and sensitive-text withholding	Detection is heuristic; do not enable translation for data that policy forbids sending externally
Host theft or malware	Restrictive OS file permissions	Enable full-disk encryption, strong login, patches, endpoint protection and encrypted backups

SECTION 13

13. Customer security checklist

- Enable the operating system's full-disk encryption (for example FileVault, BitLocker or the appropriate Linux encrypted volume).
- Use a dedicated, patched customer computer or protected server account for SM Crawler.
- Give SM Crawler access only to named, authorized users; disable accounts immediately when access is no longer required.
- Replace temporary passwords at first sign-in and use long, unique credentials.
- Protect administrator accounts and the browser profiles of paired devices.
- Grant every platform connector the minimum permissions required and revoke unused tokens at the platform.
- Keep translation disabled unless an approved provider and data-processing basis have been reviewed.
- Protect backup archives as sensitive customer data, test recovery and securely delete expired copies.
- Monitor local engine and gateway logs for failed sign-ins, connector errors and unexpected restarts.
- Keep the host running only when collection/remote access is required; understand that v194 is host-dependent.
- Do not represent v194 as end-to-end encrypted, encrypted at rest, independently penetration tested, certified or independently always-on.

SECTION 14

14. Verification evidence

This review is based on direct inspection of the v194 source and executable regression tests, including:

- **server/gateway.js**: script password verification, HMAC-signed sessions, throttling, owner/admin restrictions, tenant selection, private file modes and loopback child engines.
- **server/server.js**: local pairing, single-use codes, cookie flags, same-origin enforcement, public-file allowlist, secret-redacted configuration responses, private file writes and webhook checks.
- **server/login-privacy.js**: nine-language sign-in assurance that describes the local/central boundary.
- **server/translation.js**: consent gate, provider checks, sensitive-text withholding, private translation cache and deletion.
- **server/tunnel.js**: private token file and token-file process invocation.
- **server/connectors/catalog.js** and **server/connectors/runtime.js**: capability/visibility rules, restricted view-only routes and provider signature checks.
- **tools/make-installers.py**: explicit release allowlist, private-state exclusions, verified rollback creation and installer payload SHA-256 comparison.
- Security regression tests: remote pairing, secret redaction, origin rejection, malformed-data preservation, cookie attributes, webhook authentication, workspace separation, WebSocket isolation, tunnel-token handling and translation privacy.

The evidence is internal implementation and automated-test evidence. No independent penetration-test report, SOC 2 report, ISO 27001 certification, encryption-at-rest assessment or legal compliance opinion was found or is claimed by this document.

SECTION 15

15. Approved client-facing assurance

Your messages remain under your control. SM Crawler v194 keeps its working message history inside the customer-controlled installation and does not copy that content into a central SM Crawler customer-content database. The customer controls which SM Crawler users and operating-system administrators can access the computer and its backups. The original platforms retain their own copies, and customer-enabled external services receive data only for the function the customer chooses to activate. Local files use restricted operating-system permissions; v194 does not yet provide application-level encryption at rest or end-to-end encryption.

SECTION 16

16. Document control and release boundary

This document describes the inspected v194 implementation as of 20 August 2026. It must be reviewed again whenever authentication, storage, connector, translation, remote-access, installer, backup or deletion behavior changes. Marketing text must not shorten the assurance into “SM Crawler stores no messages,” because that would be false: the system deliberately stores working history locally. The accurate promise is no central SM Crawler customer-message database, with local customer-controlled retention and the disclosed at-rest-encryption limitation.

The independent always-on private-cloud crawler remains unimplemented. It must not be presented as available until an independent service, durable store, scheduler, reconnect/catch-up behavior, isolation controls, monitoring, recovery process and acceptance evidence exist.

DAVID BORG · All rights reserved.