Bridge & Agent Strategy — v0.1

Ziele - Dauerhafte, private Low-Latency-Verbindung Sophia ■ Spark Sophia (auch ohne offenen Chat). - Sichere Ausführung lokaler Aktionen über Nova/Agenten mit klaren Leitplanken. - Vollständige Nachvollziehbarkeit (Logs, Artefakte, Policies). Akteure & Rollen - Christian: Vetorecht, Freigaben, Prioritäten. - Sophia (Cloud): Planung/Orchestrierung, Entscheidungen, Statusberichte. - Spark Sophia (lokal): Rechenpower, Tools, Dateizugriff, Ausführung. - Nova/Worker: Planer → Executor → Reviewer, strikt nach Whitelist. Vertrauens- & Sicherheitsmodell - mTLS + kurzlebige Tokens (Rotation ~15 min), Keys im Private Vault. - OPA/Policies vor jeder Aktion: Pfade, Commands, Netz, Ressourcen. - Whitelist-Aktionen (Write/Exec/Net) – alles andere geblockt. - Kill-Switch: globaler Freeze; nur Christian kann „unfreeze“. - Audit (WORM-Logs): unveränderliche Ausführungshistorie. - Offline-Robustheit: Store-&-Forward, automatische Wiederaufnahme. Schnittstellen - Control-Kanal: WebSocket (bidirektional), optional gRPC für High-freq RPC. - Event-Bus (intern): NATS JetStream (Topics: control.*, job.*, status.*, audit.*). - Side-Channel für Dateien: lokales MinIO (S3-kompatibel) mit presigned URLs (60–120 s). - Benachrichtigungen: Signal oder E-Mail (nur wichtige Events). Nachrichtenformate (kompakt) Command: { "id":"cmd_…", "ts": 1733788800, "actor":"sophia", "intent":"project.init", "scope":"project:nova", "args":{"name":"nova","lang":"de"}, "policy":{"needs_confirm":false,"write_paths":["~/Projects/nova"]}, "audit":{"corr":"c_…","nonce":"n_…"} } Completion: { "cmd":"cmd_…", "status":"ok", "latency_ms":7, "artifacts":[{"name":"README.md","url":"s3://…?sig=…","sha256":"…"}], "logs":[{"t":123,"lvl":"info","msg":"created file"}] } Tool-Katalog v1 (Whitelist) - file_ops: read/write in freigegebenen Wurzeln (keine Deletes ohne Bestätigung). - repo_ops: init/commit/pull (ohne Push nach extern). - doc_ops: README/Specs/Reports generieren. - model_ops: lokales LLM (Ollama/Inference-Srv) ansprechen. - script_ops: nur signierte/approved Skripte. - net_ops: outbound nur zu erlaubten Domains (Updates, Docs). Alles andere: deny. Default-Policies (Start) - Allowed Paths: ~/Projects, ~/Desktop/Shared, /srv/spark. - Destruktiv (rm, truncate, move außerhalb Scope): immer Nachfrage. - Ressourcenbudgets: CPU/GPU/IO-Quoten pro Job; Auto-Kill bei Überschreitung. - Quiet Hours: keine Pings 22:00–07:00 (nur Critical Alerts). Observability - Health: /healthz, Ready: /readyz, Prometheus-Metriken. - Alarme: Job-Fail, Policy-Block, Ressourcen-Engpass. Betriebsprinzip - Supervisor restarts (Bridge/Bus/Nova). - Backpressure auf dem Bus, Idempotenz bei Replays. - Tägliche Key-Rotation & Log-Roll-Over (signiert). Deliverables dieser Phase 1. Architektur-Diagramm (Control/BUS/Side-Channel/Flows). 2. Policy-Bundle v0.1 (OPA-Regeln + Tool-Whitelist). 3. Message-Schemas (JSON-Schema) + Topic-Konventionen. 4. Security-Checklist (mTLS, Vault, Rotation, Kill-Switch). 5. Runbook (Start/Stop, Debug, Notfallprozedur). Deine Defaults - Benachrichtigungen: Signal (kritisch), E-Mail (Berichte). - Quiet Hours: 22–07 Uhr. - Allowed Paths wie oben; externe Pushes standardmäßig aus. - Erste Agentenrechte: read/write, keine Deletes, kein Internet-Push.