Closed Quorum: inside the first reported autonomous AI C2 implant

PRIMARY

Publisher	Cisco Talos
Published	2026-09-22
Retrieved	2026-09-25
Cite key	ciscotalos2026closed
URL	https://blog.talosintelligence.com/the-closed-quorum-inside-the-first-reported-autonomous-ai-c2-implant/

Published Tuesday, 22 September 2026, 06:00, by Ryan Fetterman (Cisco Talos).

The claim. "CLOSEDQUORUM is, to our knowledge, the first publicly documented Windows implant to apply this model to tactical command and control (C2)."

The vote. The implant queries four commercial LLM providers - DeepSeek, Qwen, Mistral and Google Gemini - in sequence, tallies their independent verdicts and acts on the plurality. Talos: "four models are queried in sequence, their independent verdicts tallied, and the binary acts, based on their judgment."

The tie-break. "In any tie, an order of preference kicks in: DeepSeek first, then Qwen, then Mistral, then Gemini." The implementation uses a strict "<" comparison, so the first-encountered maximum wins; DeepSeek therefore carries the deciding vote in any tie. The ordering is an artefact of the comparison, not a designed quorum rule.

The artefact. A 16.4 MB 64-bit Windows executable compiled in Go. Six SHA256 hashes across development builds spanning seven days; first analysis date 17 June 2026. Re-execution at randomised 5-15 minute intervals.

Capabilities. LSASS credential dumping; browser passwords (Chrome, Edge, Firefox); cryptocurrency wallets (MetaMask, Exodus, Ethereum); process injection and persistence. Stolen data is AES-256-GCM encrypted and exfiltrated via Discord webhooks.

Credentials. "The public distribution build, however, contains placeholder API keys and a dummy webhook, so we did not observe a complete end-to-end execution." All LLM API credentials initialise to `dummy_api_key`. Development builds demonstrate build-time injection of provider credentials, indicating per-operator compilation.

Deployment. "While we do not have confirmation of in-the-wild deployment, artifacts from the binary were used to connect the developer to postings on criminal forums related to carding, dating back to 2025."

Detection guidance. "The most useful detection strategy is still to focus on behavioral characteristics, rather than domain blocking." Named signals: AI-provider API traffic

originating from an unexpected Windows executable; similar requests potentially sent to several model providers within a short interval; Discord webhook communication from the same process or host; repeated execution at randomized 5-15-minute intervals. Talos also published a YARA rule.