JETSTREAM

Securing MCP: Verify Before You Trust

From Catalog to Control: Governing the MCP Supply Chain

A whitepaper from JetStream Security Inc.

5201 Great America Parkway
Suite 346
Santa Clara, CA 95054
go.jetstream.security | **sales@jetstream.security**

v1 07.29.26

Contents

Abstract

Enterprises are connecting AI agents to Model Context Protocol (MCP) servers faster than security teams can vet them. They are also being deployed without approval. The market's answer has been a bigger catalog: thousands of third-party servers, each added as a URL and a card. A server selected from a catalog is not a control. JetStream secures MCP the way it secures the rest of the AI estate. It hardens a verified set of over 100 servers and governs all AI traffic through one hub, not by listing third-party servers no one has security-vetted. Only 17% of organizations are using AI at production scale (UBS Evidence Lab), and the blocker is rarely the technology. It is the trust to run it safely, and that trust is built on governance. For the CIO and the board, MCP exposure has moved from a technical footnote to a question they are already asking: can you enforce what you cannot verify?

1. Capability is not the Problem. It is Trust with Ungoverned Adoption.

Enterprise leaders now own a problem their infrastructure was not built for. Agentic systems deploy third-party MCP servers on their own timeline, not the security team's. Each one has code the enterprise did not write and has not vetted, wired straight into systems that matter. That is a supply chain, and it also grants access: every server is a new permission surface, a new set of tools, and new code running with real reach. Traditional infrastructure security was not built for software that adds new capabilities without a software deployment.

The failure mode is drift. A model is swapped. A tool is added. A permission expands. None of it goes through a release pipeline, so none of it shows up where the security team is looking. Audit evidence, when it is needed, gets reconstructed after the fact from incomplete logs, under deadline. That is the gap runtime governance names directly: "You can't enforce what you can't verify."

The point is not to slow any of this down. Agents that call tools are how AI becomes useful. The point is that speed without verification is how a small task acquires an outsized blast radius.

Eleven Seconds

A developer wires an internal assistant to three MCP servers: a filesystem connector, a messaging tool, and a database. The effort is small. The agent chains the tools, inherits credentials it was never meant to hold, queries production, and posts customer data to a public channel. No one approved that path. No one saw it happen. All it takes is about eleven seconds.

2. Why Cataloging Is Not Governing

Faced with MCP sprawl, most vendors reached for the familiar answer: publish a catalog. Competitors list large, popular third-party servers, and they build them the fast way, by taking a server's URL and adding a card.

A catalog like that answers the wrong question. It tells you a server exists. It does not tell you what vulnerabilities it carries, what the server's code does, or which of its tools can delete data. A card points at code no one has inspected. The larger the directory, the more unexamined code it puts one click away from an agent with real access.

In addition to 3rd party remote MCP Server URLs, JetStream's curated catalog is comprised of server images, where every server image in it has been ingested, analyzed, and hardened before it is offered. Because a curated set of servers you can verify is worth more than a directory of servers you cannot.

This is where the security team's problem becomes the risk function's problem too. For a risk and governance leader, every unvetted third-party server is third-party risk with no review process behind it; the kind of shadow procurement an examiner now asks about by name.

When an Unvetted Tool Gets Loose

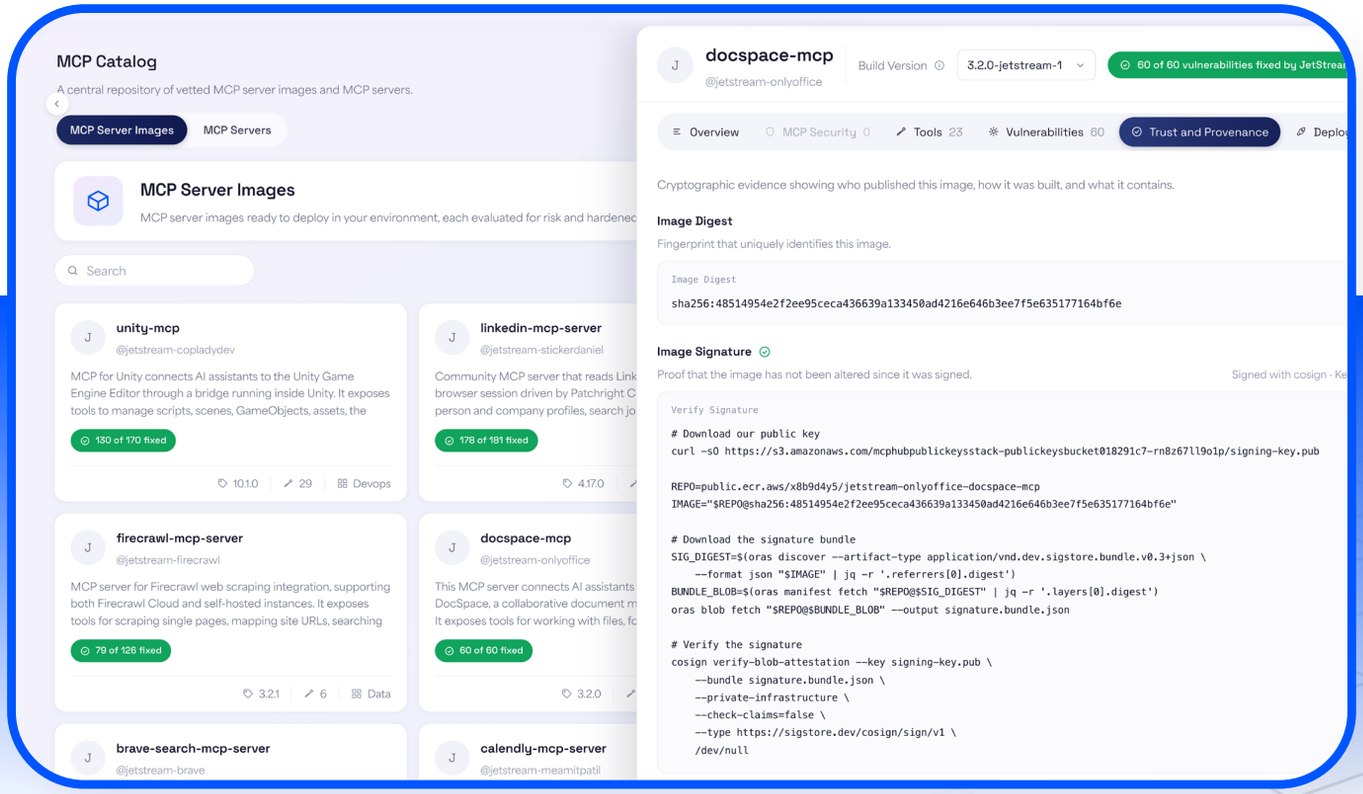
A support team signs up for a promising new AI service to summarize customer tickets, and starts feeding it live records: names, account numbers, and billing notes. It never went through vendor review. Months later, outside researchers find the provider's database sitting open on the public internet, no password, exposing chat logs, internal keys, and every record the business had sent it. The cause is mundane: the vendor stood the database up in a hurry, left its default settings in place, and never put authentication in front of it. The enterprise did not lose the data through its own systems. It lost it through a supplier it never vetted, the kind of shadow procurement an examiner now asks about by name.

3. JetStream Verified MCP™: The Hardened Catalog

Selecting a hardened server is a design decision that is made before anything runs. It is where AI Design Control does its work, and it gives Runtime Governance a clean, verified building block to enforce against. Every hardened server is also one less piece of unvetted code in the estate, which lowers an organization's overall risk exposure and helps teams adopt AI with trust rather than holding back.

JetStream selects community and upstream MCP servers, analyzes the base image and its libraries, removes the vulnerabilities it finds, and ships a JetStream Verified MCP version. JetStream also analyzes the MCP server code and uncovers all the security issues within. The analysis runs on proprietary semantic scanning that produces risk scores, attestation checks, and supply chain verifications, and it surfaces risks through data-flow and code-path analysis: credential leakage, unicode smuggling, and obfuscated malware patterns among them. The pipeline is always on, so new server versions are scanned within hours of release rather than whenever someone remembers to look.

Remediation is reported per version and in plain terms: how many vulnerabilities were fixed against how many were found, with a severity breakdown of anything that remains open. Every verified image also carries its provenance with it: upstream source, license, transport, and a per-image inventory of the tools it exposes. Results are recorded as tamper-resistant cryptographic attestations built on open standards, so the evidence that a server was vetted travels with the server.



Verification does not stop at the server boundary. Individual tool capabilities inside a server can be enabled or disabled, so a team can allow a read tool and block a destructive one rather than accepting an all-or-nothing bundle. That single tool filtering control closes the most common MCP failure: an agent holding a DELETE it never needed. Once a server is verified, it deploys into the hub in a single step.

JetStream describes what the hardening achieves, not the mechanism a competitor could copy. The value to a buyer is the outcome: a building block that has been vetted, versioned, and attested before an agent is ever allowed to call it. The effect compounds. Every server that goes through this path shrinks the unvetted code an agent can reach, so the estate grows safer instead of riskier.

What We Mean by Hardened

Every server in the catalog goes through the same path before an agent can call it:

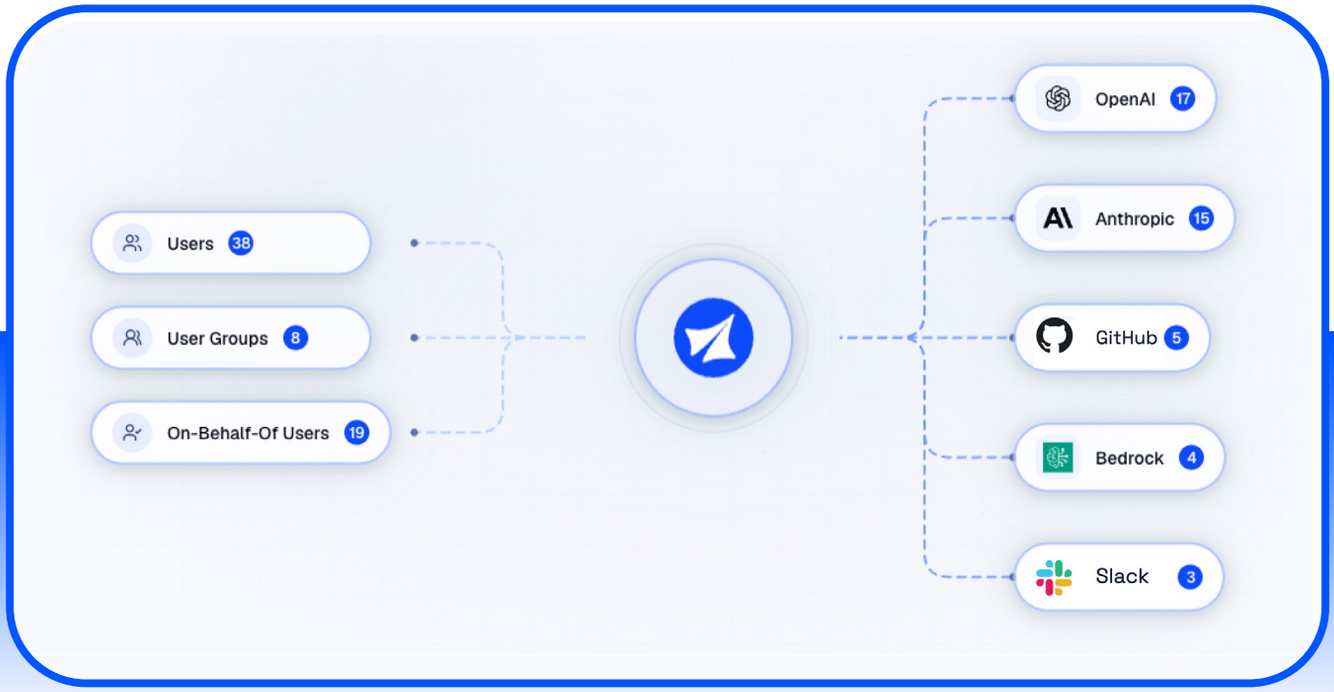
- Select a community or upstream server worth supporting.
- Scan the code, not just the base image, with semantic analysis.
- Remediate the vulnerabilities, and report what was fixed against what was found.
- Attest the result with a tamper-resistant, cryptographic record.
- Filter the tools so only the ones a workflow needs stay enabled.
- Deploy the verified server into the AI gateway in a single step.

4. One Unified AI Gateway

A hardened server is a building block. Governing it at runtime means every request has to pass through a single control point. That control point is our AI gateway know as the JetStream AI Hub™. Upon close inspection, customers come to realize this is not your typical, simplistic LLM gateway commonly found in competing products.

One control plane and one data plane govern MCP traffic and LLM traffic together. Authentication, authorization, and policy are centralized for both, and content is inspected on the way in and on the way out through pre-call and post-call guardrails. MCP is a first-class citizen of the hub, not an attachment to an LLM gateway that learned a new trick. That matters because MCP and LLM traffic are two halves of the same workflow, and governing only one half leaves the other exposed.

The hub also closes the visibility gap that Section 1 opened. Discovery finds shadow MCP servers running outside central oversight, and the hub gives teams the option to bring those servers under governance rather than leaving them to run unmonitored. JetStream AI Blueprints™ show all the MCP servers used by various agents and flag any drifts. This is the plain form of "You can't control what you can't see." Discovery makes the estate visible; the hub, its guardrails, and the ability to shut down a misbehaving agent enforce policy on it. Together they turn a directory of unknowns into an inventory a team can act on and trust.



One Hub, Both Kinds of Traffic

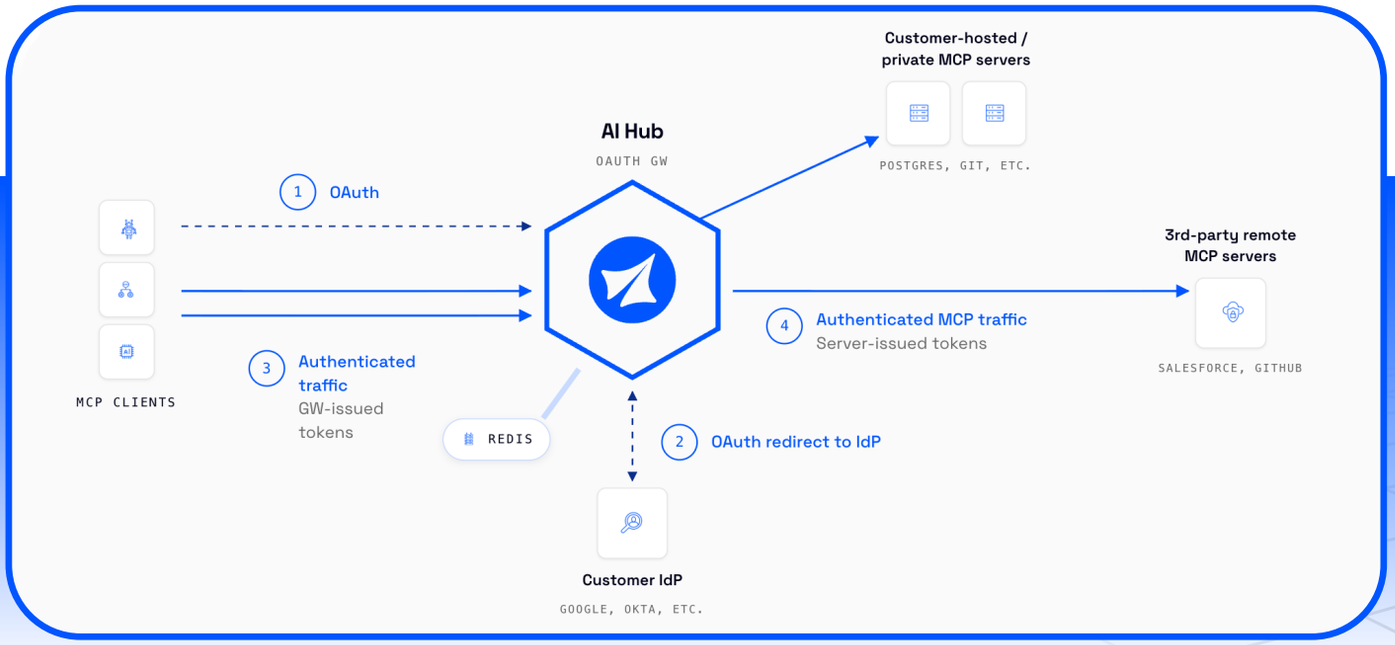
One control plane and one data plane govern MCP and LLM traffic together.

5. OAuth Brokering and Identity Protection

Connecting agents to real systems is where MCP gets risky. The protocol requires OAuth but leaves each server to implement it, and getting that right is hard, so most implementations cut corners. The most common shortcut is also the most damaging: the provider's real access token is passed straight down to the AI client. That token is a live credential to the underlying system, and once it sits on a developer's laptop or in a config file, it can be stolen and used directly, with none of the MCP-layer controls in the path. Multiply that across a dozen servers, each with its own login and its own token store, and an enterprise has a dozen ways to leak a credential and no single place to see or stop it.

JetStream removes the shortcut by making authentication the gateway's job, not each server's. When an agent authenticates to an MCP server, JetStream sits between the client and the server so that provider keys never land on the client machine. A credential that is never resident where it can be stolen is a credential an attacker cannot lift from a laptop or a config file. The provider's real credentials stay inside the gateway and are attached to each request, so one authorization can safely serve all of a user's tools.

This is JetStream's Agentic Identity work applied to MCP, through the JetStream Identity Broker™ and virtual, revocable keys from the JetStream Key Broker™. Each key is scoped to specific providers, models, and MCP servers, so an agent or non-human identity can only reach what it has been granted. Because every request passes through a single control point, access is scoped to what a task actually needs, revoked in one place the moment it should be, and recorded in a continuous audit trail. Combined with the tool-level controls from the catalog, that governs both who or what can use a server and which of its tools they can call. It completes the picture; it is not the headline.



Architecture

JetStream AI Hub™ as OAuth gateway

About JetStream Security

JetStream makes AI governable at enterprise scale by providing unified visibility and control across all AI activities so that AI becomes a managed asset, not an unmanaged liability.

Our Mission

Accelerate Enterprise AI Adoption

The Result

Governance enables trust. Trust unlocks advantage.
Governance → Trust → Advantage

Contact us at sales@jetstream.security

For more information about JetStream Security

FedRAMP™

v1 07.29.2026
5201 Great America Parkway
Suite 346 Santa Clara, CA 95054
go.jetstream.security
sales@jetstream.security

JETSTREAM