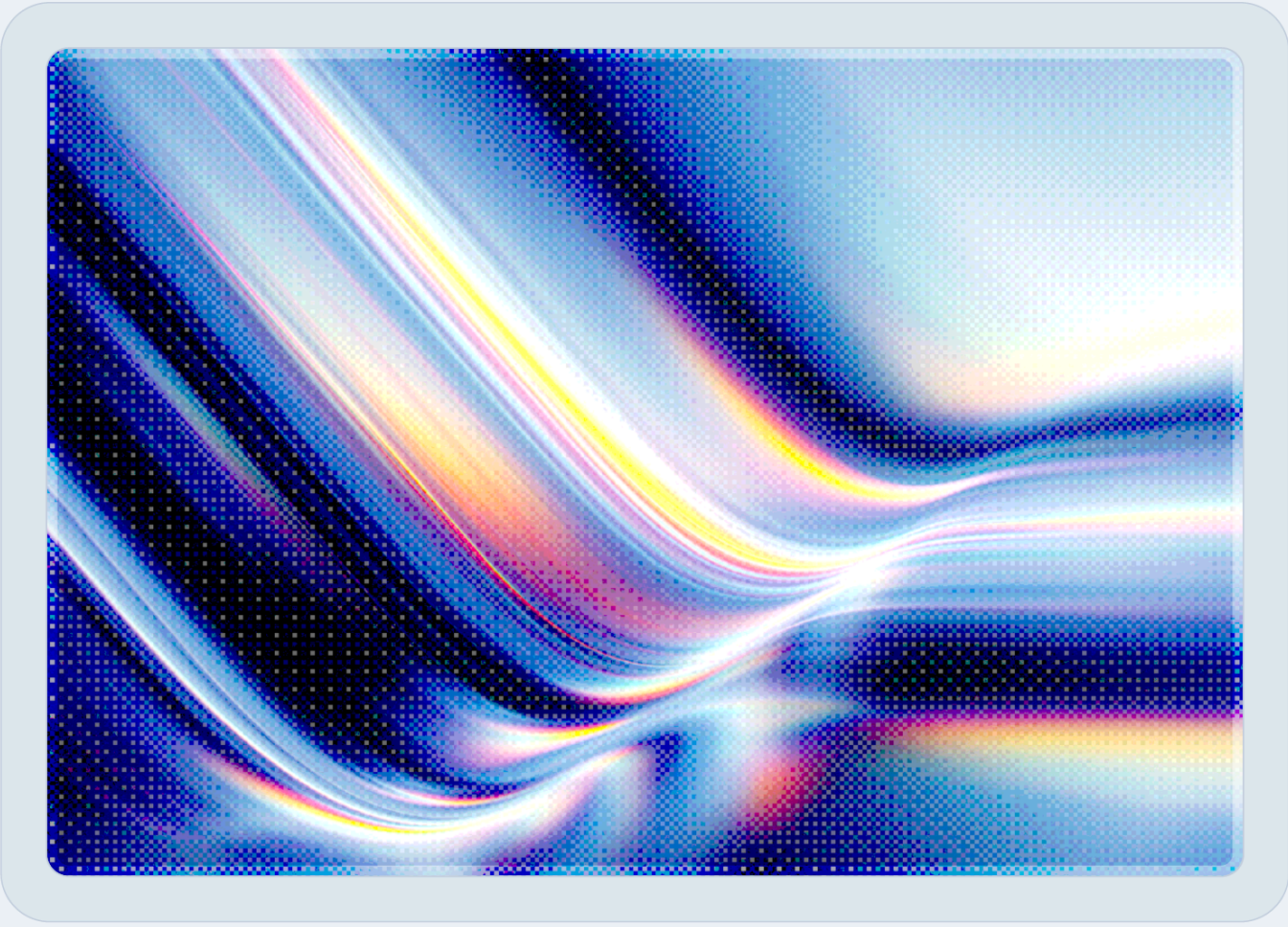
CSAI Foundation | Cloud Security Alliance

Grafana MCP Server: Session Spoofing Chained to SSRF

2026-09-03

AI-assisted Rapid Research



CSAI

CSA cloud security alliance®

Key Takeaways

Researchers at Pillar Security disclosed a critical vulnerability chain in Grafana's official Model Context Protocol (MCP) server — the tool that lets AI agents query and manage Grafana dashboards, alerts, and data sources on a user's behalf — combining a missing-authentication flaw with a server-side request forgery (SSRF) bug into a single working exploit path [1]. The SSRF component is tracked as CVE-2026-19516 and carries a CVSS score of 9.1 (CVSS:3.1/AV:N/AC:L/PR:L/UI:N/S:C/C:H/I:L/A:L), reflecting network-reachable exploitation with no user interaction required [2][3]. Pillar's research shows that, in deployments running the server without its optional authentication layer, an attacker could construct a syntactically valid but never-issued session identifier, use it to invoke the server's tools with the full authority of its configured Grafana service account, and then abuse the `grafana_api_request` tool's caller-controlled `X-Grafana-URL` header to redirect outbound requests toward internal infrastructure, including cloud metadata endpoints [1].

Grafana Labs shipped a fix in mcp-grafana v1.1.0 on August 10, 2026, adding optional bearer-token authentication for the server's network-facing transports, and the CVE was published the following day [1][2]. Because authentication remains optional rather than mandatory by default, organizations that upgrade without also enabling the new `--server-auth-token` flag remain exposed to the same failure mode the researchers identified [4]. Pillar Security estimated that the affected image had accumulated roughly 1.9 million downloads on Docker Hub prior to disclosure, indicating substantial pull volume for the affected image, though Docker Hub download counts do not distinguish unique production deployments from repeated CI/registry-mirror pulls [1]. This report explains the vulnerability chain, situates it within the broader pattern of authentication-optional MCP servers documented across 2026, and provides guidance for teams running Grafana MCP or similar internal-tool MCP integrations.

Background

The Model Context Protocol, introduced by Anthropic in late 2024, has become the de facto standard for connecting large language model agents to external tools and data sources, and Grafana Labs published an official MCP server so that AI assistants could query dashboards, inspect alerts, and pull observability data without a human operator manually navigating the Grafana UI [1]. Because the server

acts on the user's behalf using a configured Grafana service account — typically a credential with broad read and administrative access to an organization's monitoring stack — any flaw that lets an unauthorized party reach the server's tools effectively grants that party the service account's own privileges [1].

CSA's research into the broader MCP ecosystem has repeatedly found that authentication is treated as optional rather than foundational. The MCP authorization specification defines an OAuth 2.1-based framework but explicitly marks it optional for implementations, and a July 2025 internet-wide scan located more than 1,800 publicly reachable MCP servers accepting connections without any credential validation [5]. Grafana's MCP server followed this same pattern prior to version 1.1.0: it accepted session identifiers formatted like `mcp-session-<uuid>` and treated any correctly formatted string as valid, without verifying that the string had actually been issued by the server for a legitimate, authenticated caller [1]. This is functionally distinct from having no session concept at all — the server appeared to enforce sessions, which could give operators false confidence that unauthenticated access was not possible, when in practice the format check provided no security boundary [1].

This was not Grafana's first brush with SSRF risk tied to the `X-Grafana-URL` header. An earlier fix, tracked as CVE-2026-15583 and released in v0.17.2, addressed a related but narrower problem: the server was found to forward its configured Grafana credentials (service-account token, API key, or basic-auth header) to whatever host a caller specified in that header, meaning an attacker who controlled the header could exfiltrate the server's own credentials to an attacker-controlled endpoint [1][2]. Grafana's fix bound those credentials to the operator-configured Grafana URL, so a caller-specified header could no longer redirect the credentials themselves. That fix, however, did not restrict the destination of the outbound request itself — the server would still issue an HTTP request to whatever host the header named, simply without attaching its own credentials to it. Pillar Security's research shows that this residual capability was sufficient on its own to construct a working SSRF primitive once combined with the missing-authentication issue, which is the finding now tracked as CVE-2026-19516 [1][2].

Security Analysis

The vulnerability chain rests on two independent design failures that individually limit damage but combine into a critical-severity exploit path. The first failure is authentication: prior to v1.1.0, the server had no mechanism to verify that a caller presenting a session identifier was the same party the server had originally issued that session to, nor any requirement that a session be issued through an authenticated handshake in the first place [1]. An attacker who could reach the server's network endpoint — whether because it was exposed to the internet, reachable from a shared internal network, or

accessible from a compromised adjacent workload — could call the standard MCP `tools/list` and `tools/call` methods using a self-generated session identifier and have those calls processed with the server's full configured privileges [1].

The second failure is the SSRF condition in the `grafana_api_request` tool, which is designed to let an authenticated agent make arbitrary calls against the Grafana HTTP API on the user's behalf. The tool accepts caller-supplied values for the HTTP method, request path, request body, and — critically — a destination override via the `X-Grafana-URL` header, with no allowlist or validation restricting that destination to the operator's intended Grafana instance [1][2][3]. Grafana's official advisory frames the resulting condition plainly: "the caller can direct requests at internal, loopback, and link-local network services... resulting in server-side request forgery" [3]. Chained with the authentication gap, an attacker did not need any legitimate credential to reach this capability — only a correctly formatted, never-issued session identifier.

The practical consequence is that the MCP server functioned as what Pillar Security's researchers described as "a readable, method-capable proxy" positioned wherever the server itself was deployed on the network [1]. That position matters because MCP servers are frequently deployed close to sensitive infrastructure — inside a cloud VPC, alongside CI/CD systems, or within the same network segment as internal APIs — precisely because they need reachability to the tools they broker access to. An attacker exploiting this chain could direct the server to issue requests against cloud provider metadata services (a well-established path to short-lived cloud credentials), internal administrative interfaces not otherwise exposed externally, or other loopback and link-local services that assume network position alone confers trust [1][3]. The August 10, 2026 patch prevented the earlier credential-forwarding variant covered by CVE-2026-15583 from recurring, but as the vulnerability record shows, restricting where credentials travel is a different control from restricting where requests can be sent, and closing only the first gap left the second exploitable on its own [2][3].

The table below summarizes the two related CVEs in this disclosure and their current remediation status.

CVE	Root cause	CVSS	Status
CVE-2026-15583	`X-Grafana-URL` header could redirect the server's own service-account credentials to a caller-specified host	8.6 (High)	Fixed in v0.17.2

CVE	Root cause	CVSS	Status
CVE-2026-19516	Missing session authentication combined with unrestricted SSRF via `grafana_api_request` and `X-Grafana-URL`	9.1 (Critical)	Fixed in v1.1.0 (optional bearer-token auth)

Pillar Security reported both findings to Grafana through Intigriti's bug bounty program on August 2, 2026; Grafana accepted the reports and released the v1.1.0 fix on August 10, 2026 — six working days later; CVE-2026-19516 was published on August 11, 2026, and the researchers were credited in Grafana's Security Hall of Fame on August 12, 2026 [1]. This is a fast response for a design-level flaw that had been present since the server's earlier releases, though the underlying gap's persistence across prior releases is worth noting independently of the remediation speed.

Recommendations

Immediate Actions

Organizations running any version of Grafana's MCP server should upgrade to v1.1.0 or later immediately and verify the upgrade by checking the running server's reported version, since the vulnerable versions pattern-matched session identifiers without validating their issuance [1][3]. Upgrading alone is insufficient: because bearer-token authentication remains optional, operators must explicitly enable it by launching the server with the `--server-auth-token` flag and distributing that token only to legitimate callers, after which unauthenticated requests are rejected with an HTTP 401 response before any tool executes [1][4]. Teams should also audit network exposure of any MCP server, not limited to Grafana's, to confirm it is not reachable from untrusted network segments or the public internet, since the SSRF component of this chain depends entirely on the server's own network position.

Short-Term Mitigations

Where the underlying MCP server or proxy supports it, operators should implement destination allowlisting for any tool that accepts a caller-controlled URL or host parameter, restricting outbound requests to the specific, operator-configured Grafana instance rather than trusting caller input [3]. Cloud-hosted deployments should additionally block or restrict access to instance metadata endpoints from the network segment or container in which the MCP server runs, using provider-native controls

such as IMDSv2 enforcement on AWS or metadata server firewalling on other clouds, so that even a successful SSRF cannot reach credential-issuing endpoints. Security teams should also review logs for the new security warnings Grafana's v1.1.0 release introduces for unauthenticated deployments listening beyond loopback addresses, since Grafana has indicated a future release may convert that condition into a hard startup failure [3][4].

Strategic Considerations

This disclosure fits a documented pattern across the MCP ecosystem in 2026, in which the protocol's specification treats authentication as optional and individual server implementations default to the permissive option, leaving thousands of publicly reachable servers granting tool access without any credential check [5]. Organizations building or procuring MCP servers for internal tooling — observability platforms, ticketing systems, cloud control planes — should treat "authentication optional" as a red flag during security review rather than an acceptable default, and should require any internally deployed MCP server to enforce authenticated, audience-scoped access before it is connected to a production credential. Teams should also recognize that SSRF risk in MCP tools is not unique to Grafana: any tool designed to make outbound HTTP calls on a caller's behalf is a candidate for the same caller-controlled-destination flaw, and security reviews of MCP tool catalogs should specifically test for destination validation on every network-calling tool, not just assume credential handling alone is sufficient.

CSA Resource Alignment

This disclosure maps directly onto authentication and design gaps CSA has already documented across the MCP ecosystem, and the recommendations above extend that existing analysis rather than introducing a new framework.

CSA Labs' "MCP Security Crisis: Systemic Design Flaws in AI Agent Infrastructure" documents the exact structural gap underlying this vulnerability at the protocol level, noting that "the MCP authorization specification defines an OAuth 2.1 framework but explicitly marks authorization optional for implementations," and that a 2025 internet-wide scan found more than 1,800 MCP servers accepting unauthenticated connections [5]. Grafana's pre-1.1.0 behavior — accepting a correctly formatted but never-issued session identifier — is a variant of precisely this pattern, and this incident confirms that optional authentication in mainstream, widely deployed MCP servers continues to produce exploitable outcomes well after the gap was first documented.

CSA Labs' "Agentic MCP Security Best Practices Guide" is directly relevant to the remediation path Grafana chose: the guide mandates that "all remote MCP server connections must use OAuth 2.1... with PKCE mandatory for all public clients" and specifies that access tokens should be short-lived with rotation on refresh [7]. Grafana's v1.1.0 bearer-token mechanism is a meaningful improvement over the prior unauthenticated default, but it falls short of the guide's OAuth 2.1 baseline, and organizations enabling it should treat the static bearer token as an interim control rather than a substitute for the token-exchange and lifecycle practices the guide describes.

This pattern also recurs outside the MCP ecosystem narrowly defined. CSA Labs' "MLflow SSRF Under Active Attack: A Platform Breach Vector" documents an unauthenticated SSRF (CVE-2026-64849) in another AI/MLOps platform, published just eleven days before Grafana's own fix, in which the underlying condition was "validation logic that can be bypassed by a redirect, paired with a metadata service that is reachable by default." As that note observes, "the blast radius depends less on the bug itself than on what the underlying compute identity can do" — the same dynamic driving the metadata-endpoint risk described here, and a concrete precedent for treating destination validation as a standing requirement across AI tooling generally rather than a Grafana-specific fix [8].

Finally, CSA's AI Controls Matrix (AICM) v1.1 provides the control vocabulary — spanning the Identity and Access Management and Threat and Vulnerability Management domains — that organizations can use to formalize authentication and destination-validation requirements for internally deployed MCP servers as a standing procurement and architecture-review control, rather than responding to each disclosure individually [6].

References

[1] Pillar Security. "Valid But Never Issued: Session Spoofing and SSRF in Grafana MCP." Pillar Security Blog, August 2026.

[2] Vulmon. "CVE-2026-19516 - Server-Side Request Forgery in Grafana MCP Server." Vulmon, August 2026.

[3] Grafana Labs. "Security Advisories." Grafana Labs, August 2026.

[4] Grafana Labs. "mcp-grafana: MCP Server for Grafana." GitHub, 2026.

[5] Cloud Security Alliance Labs. "MCP Security Crisis: Systemic Design Flaws in AI Agent Infrastructure." CSA Labs, May 2026.

[6] Cloud Security Alliance. "AI Controls Matrix v1.1." Cloud Security Alliance, June 2026.

[7] Cloud Security Alliance Labs. "Agentic MCP Security Best Practices Guide." CSA Labs, March 2026.

[8] Cloud Security Alliance Labs. "MLflow SSRF Under Active Attack: A Platform Breach Vector." CSA Labs, August 2026.

