WHITE PAPER

Trust No Agent: Runtime Authorization for Agent-to-Agent and MCP Communications on the Fortinet Security Fabric

How enterprises discover, score, route, enforce, and audit agentic AI traffic with the Backflipt Lumen A2A Gateway and Lumen POET, on the Fortinet Security Fabric they already run, using a methodology advocated by IBM Consulting.

Backflipt FORTINET IBM

FORTINET FABRIC-READY

Backflipt is a Fortinet Fabric-Ready Technology Alliance Partner; the Lumen A2A Gateway integration is Fabric-Ready validated.

Audience: CISOs, CSOs, CIOs, Chief AI Officers, security architects **Version:** 1.0, July 2026

Publisher: Backflipt (Xenovus, Inc.), Santa Clara, CA

EXECUTIVE SUMMARY

Agents Became the Execution Layer Before Security Could Govern Them

AI has moved from chat to action. Agents now invoke tools, call APIs, exchange context, and make decisions at machine speed. Two open protocols made this shift practical at enterprise scale: the Model Context Protocol (MCP), which standardizes how agents plug into enterprise tools and data sources [3], and the Agent2Agent (A2A) protocol, which standardizes how specialist agents collaborate and delegate tasks [4]. Interoperability arrived quickly. Governance did not.

The gap is measurable. The IBM Institute for Business Value found that 82 percent of executives say secure and trustworthy AI is essential to the success of their business, yet only 24 percent of current generative AI projects are being secured [1]. Security models built for users, devices, and applications cannot see autonomous agents acting across all three. The attack surface now includes the agent communication layer itself: MCP servers, agent gateways, tool calls, and delegated access.

This white paper describes a joint architecture that places one governed control point at that layer. The Fortinet Security Fabric that enterprises already operate becomes the AI enforcement plane. The Backflipt Lumen A2A Gateway and Lumen POET add the runtime decision plane, serving as the Policy Decision Point (PDP) and Policy Enforcement Point (PEP) for both the intent of every agent call and the credentials it uses. IBM Consulting advocates the governance methodology and operates it at scale. The integration is Fabric-Ready validated as part of the Fortinet Open Ecosystem.

No agent is re-coded. No application authorization is reworked. Enterprises scale AI on the identity and security investments they already have.

The architecture is grounded in open standards throughout: MCP [3] and A2A [4] on the protocol layer, ICAP (RFC 3507) for inline traffic offload [5], OAuth 2.0 Token Exchange (RFC 8693) for credential brokering [6], SPIFFE for workload identity [7], and the zero trust principles of NIST SP 800-207 [8]. Threat coverage is mapped to the OWASP GenAI Security Project's agentic threat taxonomy, including intent breaking and goal manipulation, tool misuse, and identity and privilege abuse [2].

THE PROBLEM

The New Trust Boundary Is the Agent Communication Layer

Enterprise security control planes were built around three identities: the user, the device, and the application. Agentic AI introduces a fourth actor that traverses all three without belonging to any of them. An agent authenticates like a workload, acts on behalf of a user, calls applications like a client, and delegates work to other agents like a service. Existing controls on either side of that exchange remain effective for what they were built to do. The exchange itself, carried over A2A, MCP, tool calls, and delegated access, crosses a boundary that no existing control was designed to mediate.

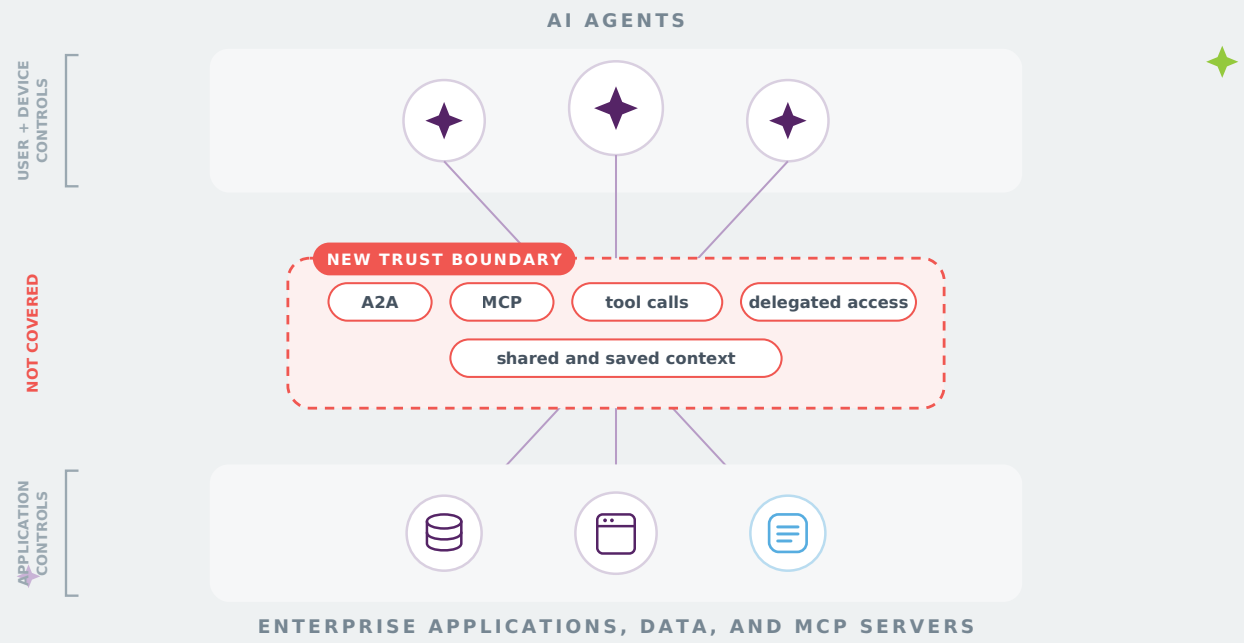

Figure 1. Agent-to-agent and MCP communications create control points that traditional user, device, and application controls were never built to govern.

Shadow AI is the new shadow IT. Autonomous agents and MCP servers appear in production faster than governance teams can manually catalog them. Most of their interactions are not authorized, monitored, or auditable, which leaves security teams unable to say yes to AI adoption with confidence. The OWASP GenAI Security Project's Agentic Security Initiative documents the resulting threat classes in its Agentic AI Threats and Mitigations taxonomy, including intent breaking and goal manipulation, tool misuse, and identity and privilege abuse [2]. These are precisely the behaviors that runtime authorization at the communication layer is designed to detect and stop.

Three Gaps Decide Whether Agentic AI Can Scale

1. Undiscovered agents

Agents call production systems before anyone registers them, creating an invisible, expanding attack surface. Governance cannot begin for an agent nobody knows exists.

2. Standing, over-scoped access

Enterprise tokens grant broad, long-lived access scoped for applications, not agents. A valid credential inside any agent quietly reaches enterprise applications, and legitimate access enables exfiltration.

3. No safe way to scale

Target applications issue permanent tokens that most of them cannot rotate. Every newly governed agent multiplies the standing credentials in circulation.

Close all three gaps and agentic AI scales safely. Leave one open and it will not.

THE METHODOLOGY

Know It, Fence It, Prove It, as One Continuous Loop

IBM Consulting advocates a four-phase governance methodology for enterprise AI, applied here to the agent communication layer and operated as a continuous loop rather than a one-time project.

82%
of executives say secure, trustworthy AI is essential to the success of their business

24%
of current generative AI projects are being secured today

Source: IBM Institute for Business Value, "Securing generative AI: What matters now," a study of C-suite executives conducted with Oxford Economics [1].

Know and Fence

Discover all AI in the network, including shadow agents, then put boundaries up so nothing reaches a resource unmediated.

Register and Attribute

Admit and allow-list agents missing from IAM, then establish who, and which agent, is accessing each resource.

Define and Enforce

Set policy decision points by user, role, and intent, then apply enforcement and swap standing tokens for short-term access.

Report

Produce detailed, framework-aligned records of enterprise AI activity for compliance and audit.

Within the joint solution, the responsibilities divide cleanly: **Fortinet identifies and steers AI traffic. Backflipt Lumen governs policy and access. IBM validates and operates at scale**, providing monitoring, evidence, and managed response so governance runs as a continuous operation.

REFERENCE ARCHITECTURE

One Inline Control Point at the Agentic Communication Layer

The reference architecture places a single governed path between every AI agent and every enterprise resource: discover, score, route, enforce, and audit. Enforcement runs on the Fortinet Security Fabric that enterprises already operate. Decisions run on the Backflipt Lumen control plane. Nothing behind the agents is re-architected.

> **The Fortinet products enterprises already operate become the AI enforcement plane. Backflipt Lumen adds the decision plane. IBM validates and operates it at scale.**

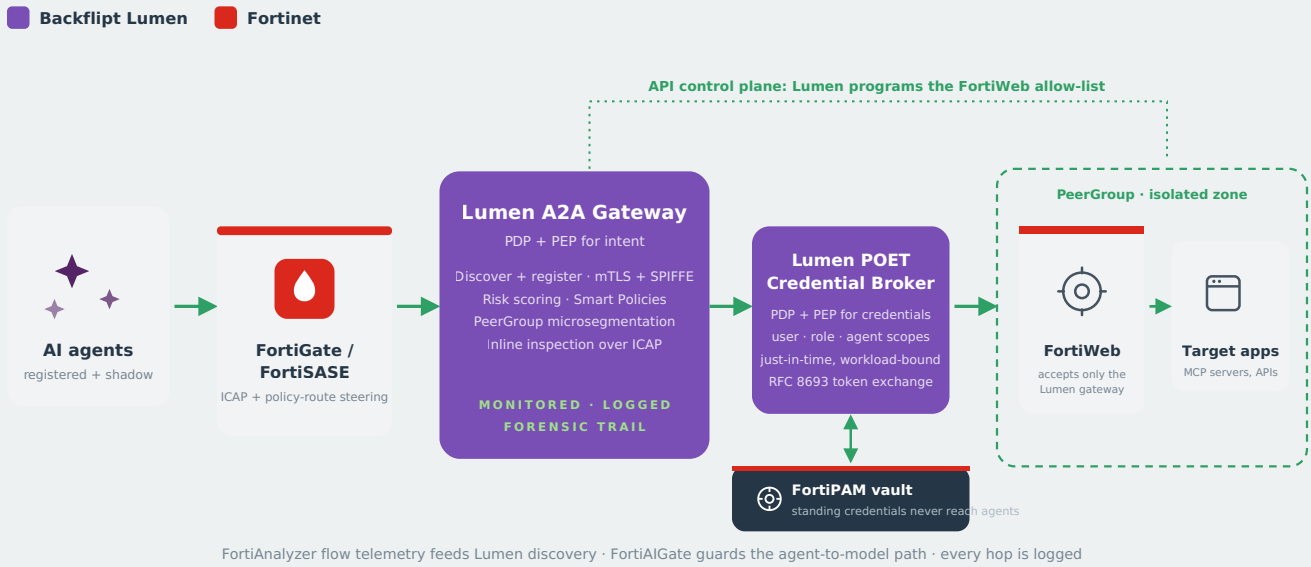

Figure 2. One governed path from agent to resource: FortiGate steers, the Lumen A2A Gateway decides on intent, the Lumen POET Credential Broker decides on credentials against the FortiPAM vault, and FortiWeb admits only gateway-mediated traffic.

Component Roles in the Governed Agent Path

Component	Role today	Role in the joint solution
FortiGate / FortiSASE	Network firewall and secure edge; ICAP offload to inspection servers	Steers agent traffic to the Lumen A2A Gateway over ICAP (RFC 3507) [5] or policy-based routing, and enforces at every boundary, fail-closed
FortiAnalyzer	Flow telemetry, logging, analytics	Supplies the network and application signals the gateway uses to discover agents and MCP servers, including shadow AI
FortiWeb	Web application firewall fronting critical targets	MCP-aware reverse proxy fronting each MCP server; accepts traffic only from the gateway
FortiAIGate	AI firewall for model traffic	Inline LLM-egress guardrail for prompt-injection, data-exfiltration, and output checks on the agent-to-model path
FortiPAM	Privileged access and secrets vault	Holds tokens, certificates, and keys out of agent reach, with policy-driven rotation, behind the Lumen POET Credential Broker

What the joint solution adds: the Backflipt Lumen A2A Gateway (runtime decision plane for intent), the Lumen POET Credential Broker (runtime decision plane for credentials), and Lumen POET (turns existing APIs into governed MCP servers and A2A agents).

RUNTIME AUTHORIZATION IN DEPTH

Lumen A2A Gateway: The PDP and PEP for Intent

The Lumen A2A Gateway is the runtime control plane for agent-to-agent and MCP traffic. It serves as both the Policy Decision Point and the Policy Enforcement Point for the **intent** of every call, evaluated through Smart Policies attached to each PeerGroup. Its capabilities map directly to the first three phases of the governance methodology.

Discovery with risk scores

The gateway discovers AI agents and the enterprise resources they call from network and application signals, including FortiAnalyzer flow telemetry, and risk-scores every finding. Administrators work a ranked queue rather than an unbounded inventory, moving each resource and its calling agents into the right PeerGroup.

PeerGroup microsegmentation

Each enterprise resource is paired with the agents approved to talk to it, default-deny for everything else. Smart Policies define which actions each PeerGroup member may take, at the granularity of individual operations, so two agents reaching the same resource can carry different effective permissions.

Inline inspection over ICAP

FortiGate hands each qualifying request to the gateway over ICAP (RFC 3507) [5], the same standards-based offload method long used for DLP and content inspection, with on-failure set to block for fail-closed enforcement. The gateway maps and checks the intent of every request before the call proceeds, addressing the OWASP agentic threat classes of intent breaking and goal manipulation, tool misuse, and goal drift [2].

Agent identity and mTLS

Agents authenticate with SPIFFE SVIDs [7], the CNCF-graduated workload identity standard referenced in NIST zero trust guidance [8, 9], or with the gateway's built-in cryptographic attestation, in both cases over mutual TLS. Identity is bound to the workload, not to a static secret.

Federated with enterprise IAM

The gateway federates with the identity infrastructure enterprises already run: bearer tokens from the enterprise IdP, Microsoft Entra agent identity blueprints, and standards-based token exchange under RFC 8693 [6]. No parallel identity silo is introduced.

First-class audit logging

Every interaction and every PeerGroup access decision is logged with the agent identity, the user it acts on behalf of, the policy applied, the scope granted, and its lifetime. The result is a forensic trail auditors can actually use.

RUNTIME AUTHORIZATION IN DEPTH

Lumen POET Credential Broker: The PDP and PEP for Credentials

Intent decisions answer whether a call should happen. Credential decisions answer what access it should carry. The Lumen POET Credential Broker, an add-on service to Lumen POET, front-ends the generated MCP servers and A2A agents and serves as the Policy Decision Point and Policy Enforcement Point for **credentials** at the last mile.

The problem it removes is structural. Enterprises secured their applications and data, then agents arrived inside that estate holding the tokens the targets issue: permanent, broad in scope, designed for fixed applications. Every governed agent adds a standing credential, the blast radius grows with the fleet, and each new static secret in circulation is a program failure, not an operating cost. Most targets cannot accept anything shorter-lived, so the fix has to happen in front of them.

The inbound order to every MCP server is fixed: broker first, FortiWeb second, server third. The Lumen A2A Gateway has already authorized the call's intent upstream, so the broker's decision is purely about access.

1. The schema

The application administrator defines the PDP: which user, role, or AI agent maps to which features and scope of the target application. Microsoft Entra agent blueprints are supported.

2. The validation

At request time the broker validates the caller with AD or Entra. In Entra estates it uses On-Behalf-Of, Microsoft's implementation of OAuth 2.0 Token Exchange (RFC 8693) [6]; the broker registers with the directory as an application and receives a time-bound token.

3. The swap

The broker exchanges the caller's token for the application's own token held in the PAM vault: OAuth, API key, or Basic Auth, whichever the target expects. The agent never sees it.

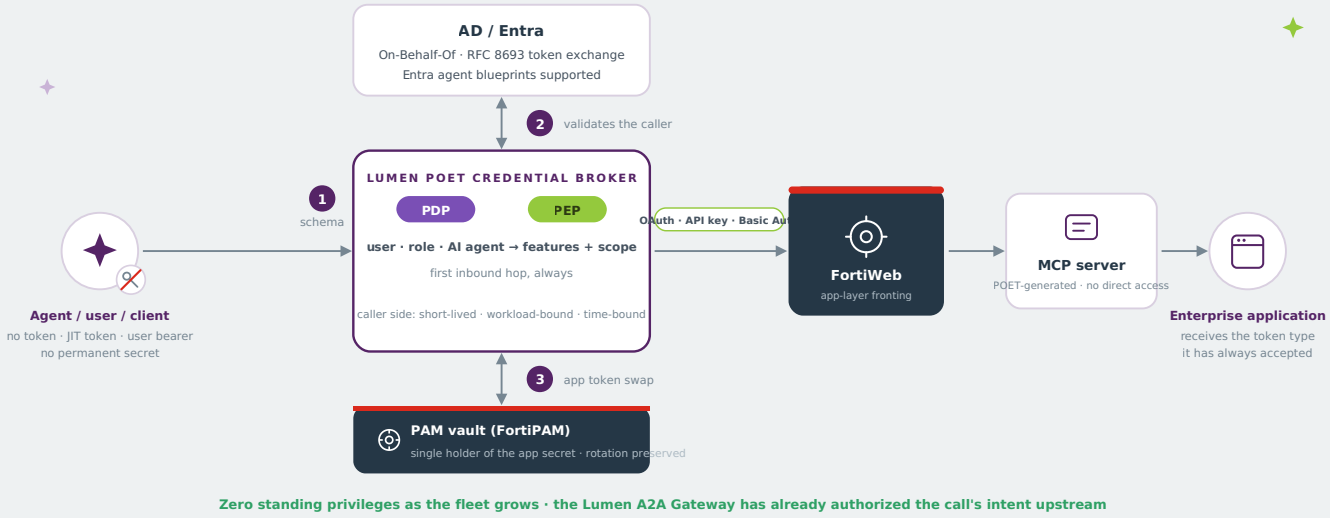

Figure 3. The fixed inbound order to every MCP server: Credential Broker first, FortiWeb second, server third. Standing secrets never leave the vault, and the caller side carries only short-lived, workload-bound, time-bound access.

What Follows

Vault rotation is preserved

The vault remains the single holder of the application secret, and the PAM team's rotation scheme keeps running exactly as it does today.

Legacy targets become compliant without a code change

Workload-based, temporary access on the caller side; the token type the application has always accepted on the target side.

Zero standing privileges across the agent fleet

A compromised agent token is task-scoped and time-bound; the application token was never on the agent.

Audit evidence per transaction

Caller, schema decision, credential type, and credential lifetime, recorded alongside the gateway's intent verdict.

> **Standing secrets live only in the vault. Agents hold nothing worth stealing.**

Lumen POET: Governed Agents from the APIs Enterprises Already Own

Lumen POET generates precise MCP servers and A2A agents directly from OpenAPI, Swagger, or RAML definitions, with no code. Generated services deploy via Helm or Terraform inside the customer's own VPC and coexist with the API gateways enterprises already run, including Kong, Apigee, AWS API Gateway, and MuleSoft. Because the broker front-ends everything POET generates, every enterprise API becomes agent-ready under the same policy regime from day one. The reference implementation runs in an IBM lab, a secure Fortinet environment where customer prototypes can be provisioned. No agent is re-coded and no application authorization is reworked.

DEPLOYMENT ENGINEERING

Traffic Steering: Getting Every Agent Call Into the Governed Path

A runtime PDP and PEP is only as strong as its insertion point. The joint solution supports three validated steering patterns on FortiGate, so that qualifying agent and MCP traffic transits the Lumen A2A Gateway without changes to the agents or the targets.

Pattern 1: ICAP offload (recommended for gateway-adjacent inspection)

FortiGate hands each HTTP request to the gateway over ICAP (RFC 3507) [5], the same offload mechanism it has long supported for DLP and content inspection. Configuring on-failure to block yields fail-closed enforcement: if the decision point is unreachable, the call does not proceed.

Pattern 2: Policy-based routing for east-west agent traffic

Where a source agent and a target service sit on different internal subnets and the routed path already transits the FortiGate, a standard interface and next-hop policy route redirects qualifying traffic inline to the gateway. Three engineering facts make this pattern robust in production:

Destination preserved

The policy route matches on Layer 3 and Layer 4 fields and sets the next hop to the gateway. The destination IP of the target is never rewritten; only the egress frame's MAC is resolved to the gateway via ordinary ARP. No static MAC mapping is required.

Termination model

The gateway terminates the connection, runs the policy decision, and re-originates to the target. Return traffic then belongs to the gateway's own session, so no return-path policy route is needed and inspection sees both directions.

Asymmetry avoided

Transparent forwarding, by contrast, lets the reply bypass the gateway unless a mirrored return-path policy route is added. The termination model is recommended for the A2A and MCP enforcement point precisely because it removes this trap.

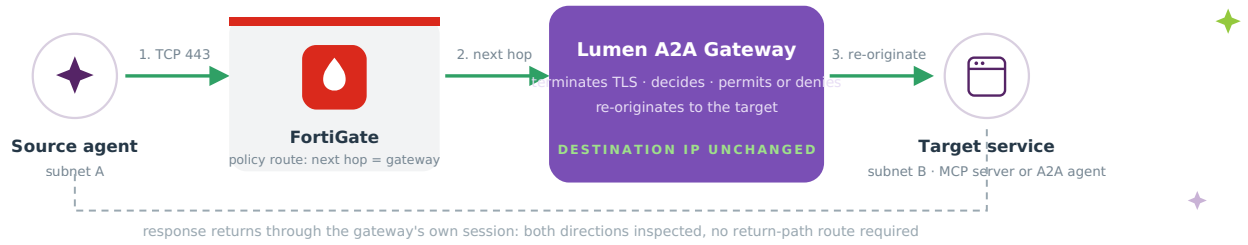

Figure 4. East-west insertion with a FortiGate policy route: the destination IP is preserved end to end, and the termination model keeps both directions of every exchange inside the inspection path.

Pattern 3: Virtual IP with destination NAT

Where the gateway already holds the upstream mapping, a FortiGate Virtual IP rewrites the destination to the gateway and the gateway operates as an explicit reverse proxy with the target configured as its upstream. This

variant is operationally simple and avoids transparent-proxy requirements on the gateway host, at the cost of true transparency to the source.

Full configuration detail for the policy-route pattern, including FortiOS objects, packet processing order, and verification commands, is available from Backflipt in the deployment note "Steering A2A and MCP traffic to the Lumen A2A Gateway with FortiGate policy-based routing" [10].

STANDARDS ALIGNMENT

Built on Open Standards End to End

Every interface in the joint architecture is standards-based. Enterprises evaluating the solution can validate each claim against the primary source.

Standard	Steward and status	Role in the architecture
MCP	Open protocol introduced by Anthropic, November 2024; open specification and community governance [3]	The tool and data integration protocol the gateway inspects and POET generates servers for
A2A	Open protocol contributed by Google to the Linux Foundation, June 2025 [4]	The agent collaboration protocol governed per PeerGroup by the gateway
ICAP, RFC 3507	IETF Informational RFC, April 2003 [5]	Standards-based inline offload from FortiGate to the gateway, fail-closed
OAuth 2.0 Token Exchange, RFC 8693	IETF Standards Track RFC, January 2020 [6]	Token exchange in the gateway's IAM federation and the broker's credential exchange
SPIFFE	CNCF graduated project [7]	Workload identity for agents via SVIDs over mutual TLS
NIST SP 800-207 / 800-207A	NIST Special Publications on zero trust architecture [8, 9]	The zero trust principles the architecture applies to non-human, agentic actors; SP 800-207A explicitly identifies SPIFFE-based workload identity as a ZTA building block
OWASP Agentic AI Threats and Mitigations	OWASP GenAI Security Project,	The threat taxonomy the intent checks and

Standard	Steward and status	Role in the architecture
	Agentic Security Initiative [2]	credential controls are mapped against

Evidence for Compliance Programs

Every agent call in the governed path is logged with the agent identity, the user it acts on behalf of, the policy applied, the scope granted, and its lifetime. Framework-aligned records support programs under the EU AI Act, ISO 42001, NIST frameworks, SOC 2, HIPAA, and PCI DSS.

ADOPTION

Start with Visibility, Prove Control, Then Scale

1. Assess

Begin with a trusted AI exposure assessment to discover agentic activity, including shadow agents and unregistered MCP servers.

2. Validate

Validate priority use cases in a lab or cyber range environment. The joint environment runs today in the Fortinet Lab at IBM's Software Lab in Markham, Ontario, where financial services and government organizations model their own configurations before deployment.

3. Define

Define policies, controls, and enforcement points across MCP and agent traffic: PeerGroups, Smart Policies, and broker scopes.

4. Scale

Scale into production through managed operations with IBM Consulting: monitoring, evidence, and managed response as a continuous loop.

Backflipt, Fortinet, and IBM Consulting present **"Trust No Agent: Securing Agent-to-Agent and MCP Communications"** at Black Hat USA 2026 on Wednesday, August 5 at 2:00 PM in the Fortinet Theater, including a live demonstration of the governed agent lifecycle on an active Fortinet fabric. Fortinet ecosystem inquiries: **fabricready@fortinet.com**.

REFERENCES

Sources and Further Reading

1. IBM Institute for Business Value, with Oxford Economics, "Securing generative AI: What matters now," 2024. **https://www.ibm.com/thought-leadership/institute-business-value/report/securing-generative-ai**
2. OWASP GenAI Security Project, Agentic Security Initiative, "Agentic AI Threats and Mitigations," 2025. **https://genai.owasp.org/resource/agentic-ai-threats-and-mitigations/**
3. Model Context Protocol, specification and documentation. **https://modelcontextprotocol.io**
4. Agent2Agent (A2A) Protocol, a Linux Foundation project. **https://a2a-protocol.org**; Linux Foundation launch announcement, June 23, 2025: **linuxfoundation.org/press**
5. J. Elson and A. Cerpa, "Internet Content Adaptation Protocol (ICAP)," IETF RFC 3507, April 2003. **https://www.rfc-editor.org/rfc/rfc3507**
6. M. Jones, A. Nadalin, B. Campbell, J. Bradley, and C. Mortimore, "OAuth 2.0 Token Exchange," IETF RFC 8693, January 2020. **https://www.rfc-editor.org/rfc/rfc8693**
7. SPIFFE, the Secure Production Identity Framework for Everyone, a CNCF graduated project. **https://spiffe.io**
8. S. Rose, O. Borchert, S. Mitchell, and S. Connelly, "Zero Trust Architecture," NIST Special Publication 800-207, August 2020. **https://csrc.nist.gov/pubs/sp/800/207/final**
9. NIST, "A Zero Trust Architecture Model for Access Control in Cloud-Native Applications in Multi-Cloud Environments," NIST Special Publication 800-207A, 2023. **https://csrc.nist.gov/pubs/sp/800/207/a/final**
10. Backflipt, "Steering A2A and MCP traffic to the Lumen A2A Gateway with FortiGate policy-based routing," deployment note v1.1, 2026. Available from Backflipt at **backflipt.com/datasheets**.
11. Fortinet, Technology Alliances and the Fabric-Ready Partner Program. **https://www.fortinet.com/partners/technology-alliances/alliances-ecosystem**; Open Ecosystem partner directory: **fortinet.com/partners/partnerships/alliance-partners**
12. Fortinet, "Backflipt Alliance Solution Brief." **https://www.fortinet.com/content/dam/fortinet/assets/alliances/asb-backflipt.pdf**

All references verified as of July 2026. Statistics are quoted from their primary published sources; readers are encouraged to consult each source directly.

About Backflipt

Backflipt, a brand name of Xenovus, Inc., is an enterprise software company headquartered in Santa Clara, California. Backflipt builds the Lumen product line for securing and governing AI agents in regulated enterprises: the Lumen A2A Gateway, an inline policy decision-and-enforcement point for agent-to-agent and MCP traffic, and Lumen POET, a no-code platform for transforming APIs into governed agents and brokering ephemeral credentials. Founder-funded since 2013, Backflipt has production deployments in regulated Fortune 500 environments. For more information, visit **backflipt.com**.

About Fortinet

Founded in the San Francisco Bay Area in 2000, Fortinet continues to be a driving force in the evolution of cybersecurity and the convergence of networking and security. Securing people, devices, and data everywhere is its mission. Fortinet's portfolio of over 50 enterprise-grade products is the largest integrated offering available, delivering proven cybersecurity everywhere it is needed. More than 900,000 customers trust Fortinet solutions, which are among the most deployed, most patented, and most validated in the industry. For more information, visit **fortinet.com**.

About IBM Consulting

IBM Consulting helps enterprises adopt AI securely, advocating the trusted AI governance methodology in this solution: assessment of agentic exposure, validation of priority use cases in lab and cyber range environments, and managed operations that keep monitoring, evidence, and response running at scale. For more information, visit **ibm.com/consulting**.

Topics: #Fortinet · #FortinetFabricReady · #FabricReady · #IBM · #IBMConsulting · #IBMAI · #AgenticAI · #AIAgents · #AISecurity · #AIGovernance · #ZeroTrust · #ZTNA · #MCP · #MCPSecurity · #A2A · #NonHumanIdentity · #NHI · #ShadowAI · #CyberSecurity · #TrustNoAgent · #Backflipt · #LumenA2AGateway · #LumenPOET · #FortiGate · #EphemeralCredentials · #WorkloadIdentity · #WorkloadIAM · #MachineIdentity · #ZeroStandingPrivileges · #WhitePaper