SAWBILL

WHITE PAPER / 01

TRUST

From Declared Trust to Verifiable Authority

Identity, delegation, and decision rights in an agentic system

How Sawbill turns a verified identity into precise, limited, revocable authority bound to an action.

READERS

Architects / Security leaders / AI platform leaders / Risk and governance leaders

Contents

From Declared Trust to Verifiable Authority

The issue in one minute

When an agent presents itself as a "development agent," "reviewer," or "administrator," the label proves nothing. Even a correctly authenticated identity does not answer the decisive question: is this identity allowed to perform **this action**, on **this object**, in **this context**, at **this time**?

Sawbill replaces implicit trust with a verifiable chain:

> **a credential is verified, an identity is resolved, authority is evaluated, and its use is bound to an exact request.**

The result is neither a global role nor a reusable token. It is a bounded, traceable, and invalidatable authorization context.

Why conventional controls are no longer enough

Traditional systems often rely on four shortcuts:

- the signed-in account is assumed to represent the right person;
- the `admin` role is assumed to cover every necessary decision;
- the worker executing a task is assumed to act for the right requester;
- a valid signature is assumed to mean approval.

These shortcuts become dangerous when agents can delegate, change model or runtime, work across repositories, and request external effects. An account can be compromised. An agent can be restarted under a different workload identity. An authorization valid for a repository is not necessarily valid for production infrastructure.

Sawbill therefore treats six objects separately, even though applications often conflate them:

Object	Question it answers	What it does not prove
Credential	What authentication material is presented?	Identity or the right to act
Verified identity	Which principal is this material bound to?	A role or permission
Trust context	Which anchors, policies, and revocation data were used?	A business decision
Granted authority	Which actions are allowed on which subjects?	That the action is admissible now
Delegation	What bounded portion of parent authority is transferred?	An expansion of power
Authorization context	Is this exact use allowed?	That the action started or succeeded

This separation is the foundation of the Sawbill trust model.

The Sawbill authority chain

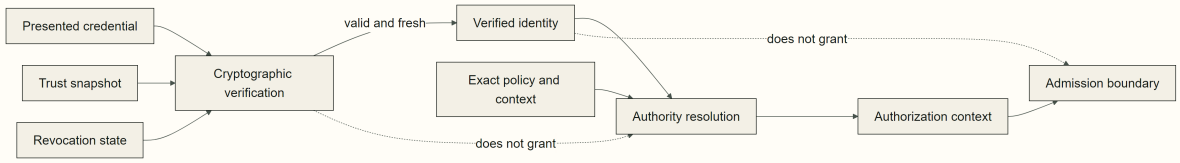

Figure 1 - Sawbill

1. Verify the credential

Sawbill first verifies the material actually presented: a certificate, identity assertion, workload credential, or another mechanism admitted by the applicable profile. Verification binds, among other things:

- the expected format and version;
- the chain and trust anchor;
- the credential's intended use;
- its validity period;
- available revocation information;
- the time and context of verification.

An unknown chain, unavailable revocation check, or ambiguous profile does not become "probably valid." The result remains blocked or unknown.

2. Resolve a stable identity

The credential is then linked to a stable Sawbill identity. That identity is not an email address, display name, or session ID. Credentials can rotate without changing the principal; conversely, two credentials carrying the same name are not automatically equivalent.

For workloads, Sawbill can rely on short-lived, rotating identities such as the SVIDs defined by SPIFFE. SPIFFE provides verifiable workload identities and trust domains. Sawbill retains the business decision that determines what the workload is allowed to do.

3. Evaluate authority

Sawbill authority describes an explicit relationship:

- the authorized principal;
- the permitted actions;
- the subjects or resources concerned;
- scope, time, or project constraints;
- delegation limits;
- the provenance of the authority;
- the policy governing its interpretation.

Resolution compares those data with the real request. It does not merely ask whether the principal "has the developer role," but whether that principal may, for example, propose a change to Project A's export service during a specific attempt.

4. Bind authorization to an exact use

The result is an authorization context bound to the action, subject, contract, attempt, audience, and decision time. Its use remains limited to the purpose for which it was produced.

Copying that context to another attempt or resource therefore fails, even when the same agent and credential are present.

Where Sawbill fits into the existing identity stack

Sawbill does not replace the directory, certification authority, or workload identity service. Those systems continue to authenticate identities in their own scope. Sawbill adds the decision missing from an agentic delivery chain: what exact action may this identity request now?

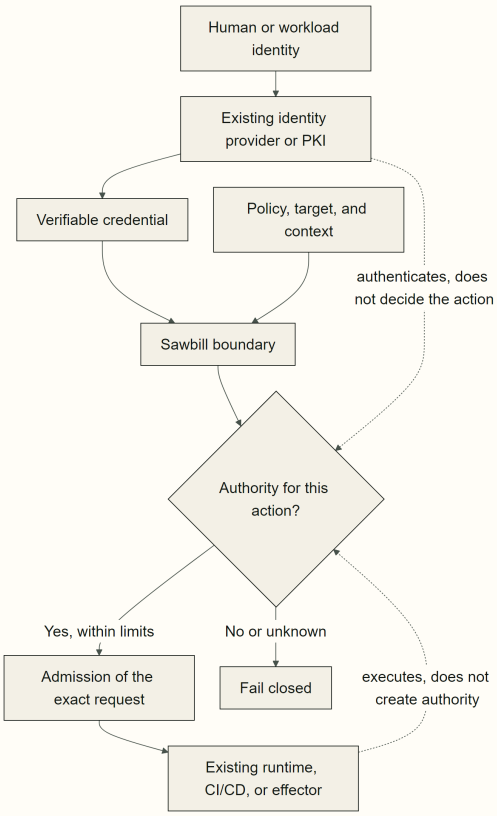

Figure 2 - Sawbill

Workload identity standards, X.509 certificates, and authentication mechanisms therefore remain integration points. They provide a verifiable identity; Sawbill retains business authority, delegation, freshness, and admission. The trust boundary moves from the account or runtime to an explicit decision bound to the request.

A certificate identifies; Sawbill authorizes

Sawbill relies on PKI without delegating all governance to it. The X.509 profile follows the certificate and path validation foundations of RFC 5280. An owner certificate can bind a public key to an identity, installation, and scope. On its own, it never becomes a right to change a baseline, accept a deliverable, or perform a deployment.

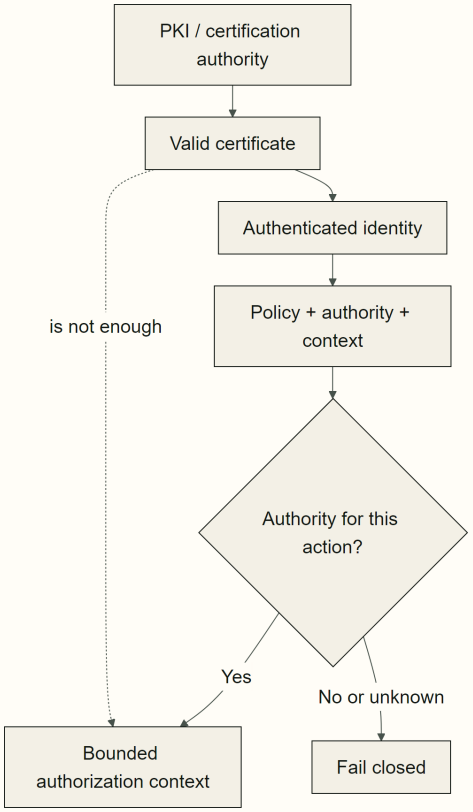

Figure 3 - Sawbill

This architecture accepts a trust chain selected and administered by the organization. The owner's private key remains under the owner's control and is not handed to agents. Issuance and renewal can remain with the existing infrastructure; the certificate service never decides Sawbill authority.

Delegate without expanding

Delegation is essential in multi-agent systems: an orchestrator assigns an analysis to a specialist, who may in turn request a verification. Naive delegation, however, quickly turns limited permission into transitive power.

Sawbill applies a simple rule: a child delegation must be a mechanically verifiable attenuation of its parent.

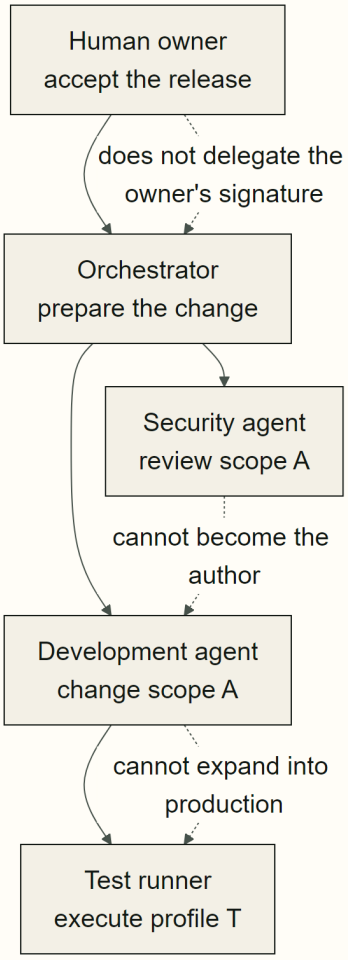

Figure 4 - Sawbill

Each child retains an explicit relationship with its parent, a depth, a scope, and constraints at least as strict. An incomparable or ambiguous value blocks delegation. Revocation or expiry of an ancestor makes its descendants unusable under the applicable policy.

Separation of duties depends on actual control

Two different agent names do not prove independence. They may share the same model, provider account, key, runtime, or human controller.

Sawbill can evaluate separation of duties across several dimensions:

- principal and controlling principal;
- organization or team;
- credential and key domain;
- workload and execution environment;
- model and model profile;
- relationship among author, validator, and approver.

Independence remains proportional to risk. A mechanical control can verify that two credentials and runtimes differ; it does not claim to prove the absence of organizational collusion without suitable information.

Scenario: authorize a data export capability

A product manager requests a customer data export capability. The work touches an API, an infrastructure repository, and a data residency policy.

1. Her credential is verified under the organization's profile.
2. Her Sawbill identity is resolved independently of her application account.
3. Her authority allows creation of the intent, but not security approval or production deployment.
4. The orchestrator receives a delegation limited to preparing the change.
5. The development agent may modify the defined scope; it cannot expand its mission to the adjacent billing system.
6. The security reviewer is selected under the required independence criteria.
7. Final approval requires the authority and, depending on the profile, the quorum specified for the exact deliverable version.

The system never relies on the idea that "the agent works for the right team." Each important transition has its own authority.

What this approach changes for the enterprise

For security

- fewer standing permissions entrusted to agents and workers;
- revocation and rotation reflected in current decisions;
- rejection of escalation based on ambiguous accounts, roles, or labels;
- verifiable separation among identity, authority, and execution.

For engineering

- multi-agent delegation without loss of traceability;
- reusable policies across interfaces and environments;
- reproducible decisions from the same inputs;
- integration of multiple identity providers without merging their semantics.

For risk and audit

- attribution of every decision to a verified principal;
- explicit scope and limits for authority;
- history of rotations, revocations, and delegations;
- clear distinction between authentication and approval.

What Sawbill establishes - and what it does not

Sawbill can establish	Sawbill does not automatically infer
A credential was verified under a precise profile and context	That every claim carried by the credential is true
An identity was resolved at a specific cutoff	That the identity has a right to act
Authority permits a bounded action	That the action should be executed or accepted

Sawbill can establish	Sawbill does not automatically infer
A delegation attenuates its parent's power	That the actors are organizationally independent in every circumstance
An authority decision was made over exact inputs	That the decision remains usable after revocation or a policy change

This precision does not weaken trust. It makes trust verifiable.

Official references

Source	External status	Contribution used	Limit retained by Sawbill
RFC 5280 - PKIX Certificate and CRL Profile	IETF Proposed Standard, updated by several RFCs	X.509 certificates, extensions, path validation, and CRLs	A certificate is not a Sawbill authorization
NIST SP 800-63-4 - Digital Identity Guidelines	Final version published in July 2025	Assurance vocabulary for identity proofing, authentication, and federation	Assurance levels do not become business permissions
SPIFFE Standard	Maintained open specification; Workload API declared stable	Portable workload identity, SVIDs, and trust domains	A SPIFFE identity authenticates a workload; Sawbill decides its authority
NIST SP 800-207 - Zero Trust Architecture	Final NIST publication	No implicit trust based on network location or ownership	Sawbill applies the principle to agentic actions and governed objects

External references remain normative within their own scope. Their inclusion is not a Sawbill certification, a legal guarantee, or equivalence with an assurance level that has not been evaluated.