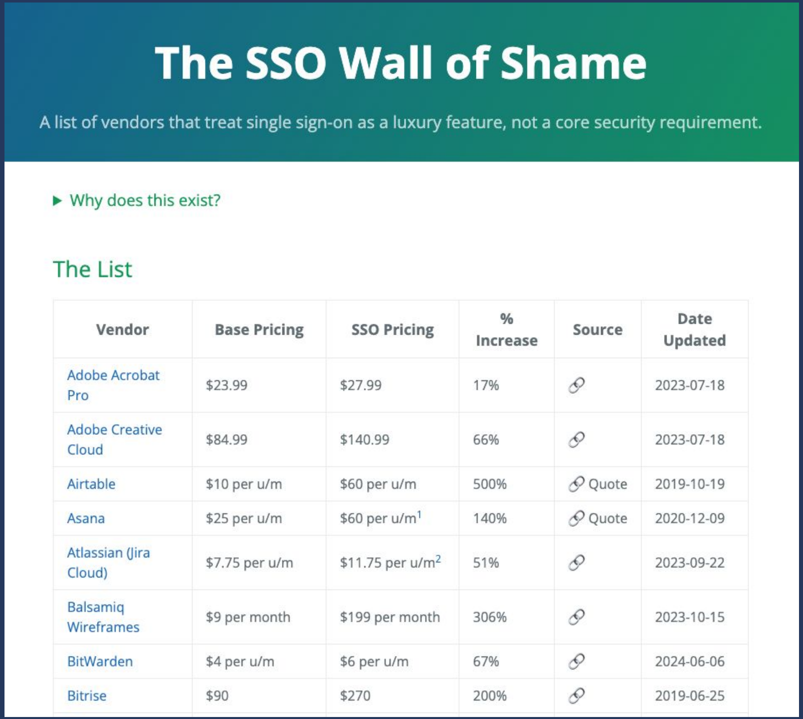
The Future of AuthZ, *from A to Z*

Dr. Rohit Khare

IRS IAM AuthZ All-hands — August 6th, 2024

AuthZ.Substack.com & DecentIAM.com

An “SSO . Tax” is Inevitable – *Finally!*



The SSO Wall of Shame

A list of vendors that treat single sign-on as a luxury feature, not a core security requirement.

▶ Why does this exist?

The List

Vendor	Base Pricing	SSO Pricing	% Increase	Source	Date Updated
Adobe Acrobat Pro	$23.99	$27.99	17%	🔗	2023-07-18
Adobe Creative Cloud	$84.99	$140.99	66%	🔗	2023-07-18
Airtable	$10 per u/m	$60 per u/m	500%	🔗 Quote	2019-10-19
Asana	$25 per u/m	$60 per u/m[1]	140%	🔗 Quote	2020-12-09
Atlassian (Jira Cloud)	$7.75 per u/m	$11.75 per u/m[2]	51%	🔗	2023-09-22
Balsamiq Wireframes	$9 per month	$199 per month	306%	🔗	2023-10-15
BitWarden	$4 per u/m	$6 per u/m	67%	🔗	2024-06-06
Bitrise	$90	$270	200%	🔗	2019-06-25

It took decades, but Authentication standards replaced “roll-your-own” user login & password systems

When will Authorization standards reach the same level of maturity?

- *Why* will we need AuthZ?
- *What* problems will it solve?
- *How* soon will it be adopted?

Outline of this talk

1. ***Who* I am,**
   and why I care about a *crisis of complexity* in Cloud security
2. ***What* I've learned about AuthZ,**
   from *editing a newsletter* about AuthZ every week in 2024
3. ***Why* I believe AuthZ must interoperate,**
   to connect *decentralized services* and enable safe AI agents

***How* I envision AuthZ will work with LLMs & automated reasoning,**
to *simplify security* using "*proven policies in plain English*"

Dr. Rohit Khare, Software Architect

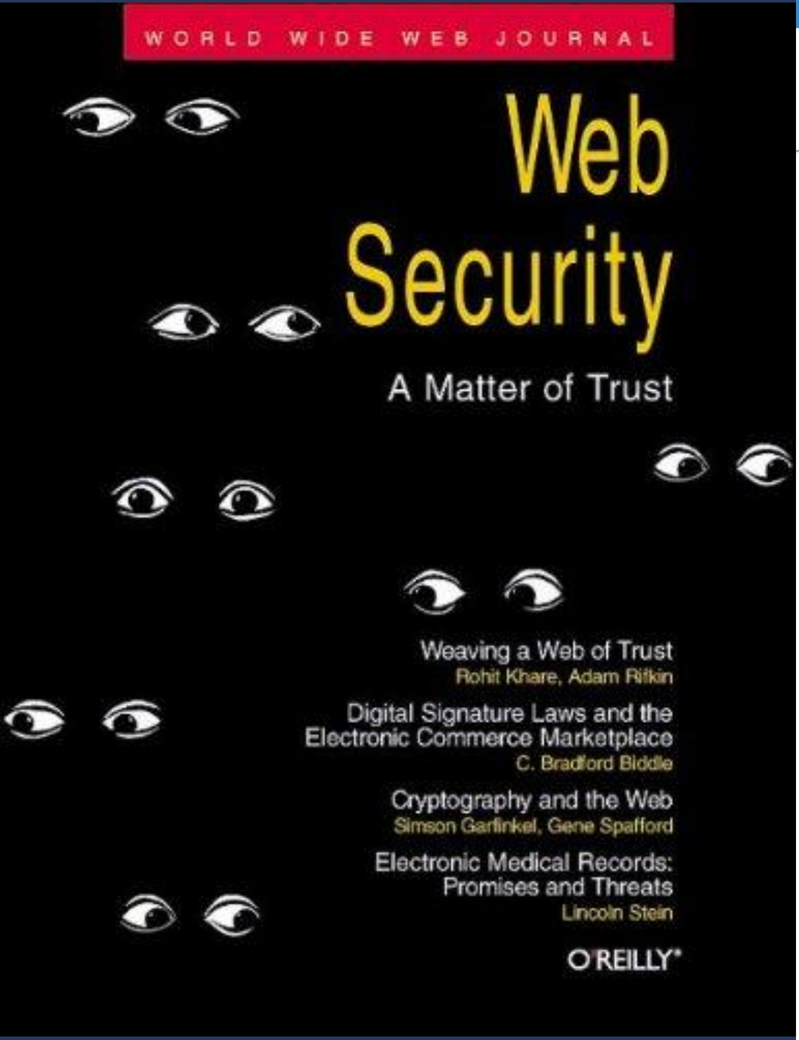
“Forrest Gump”

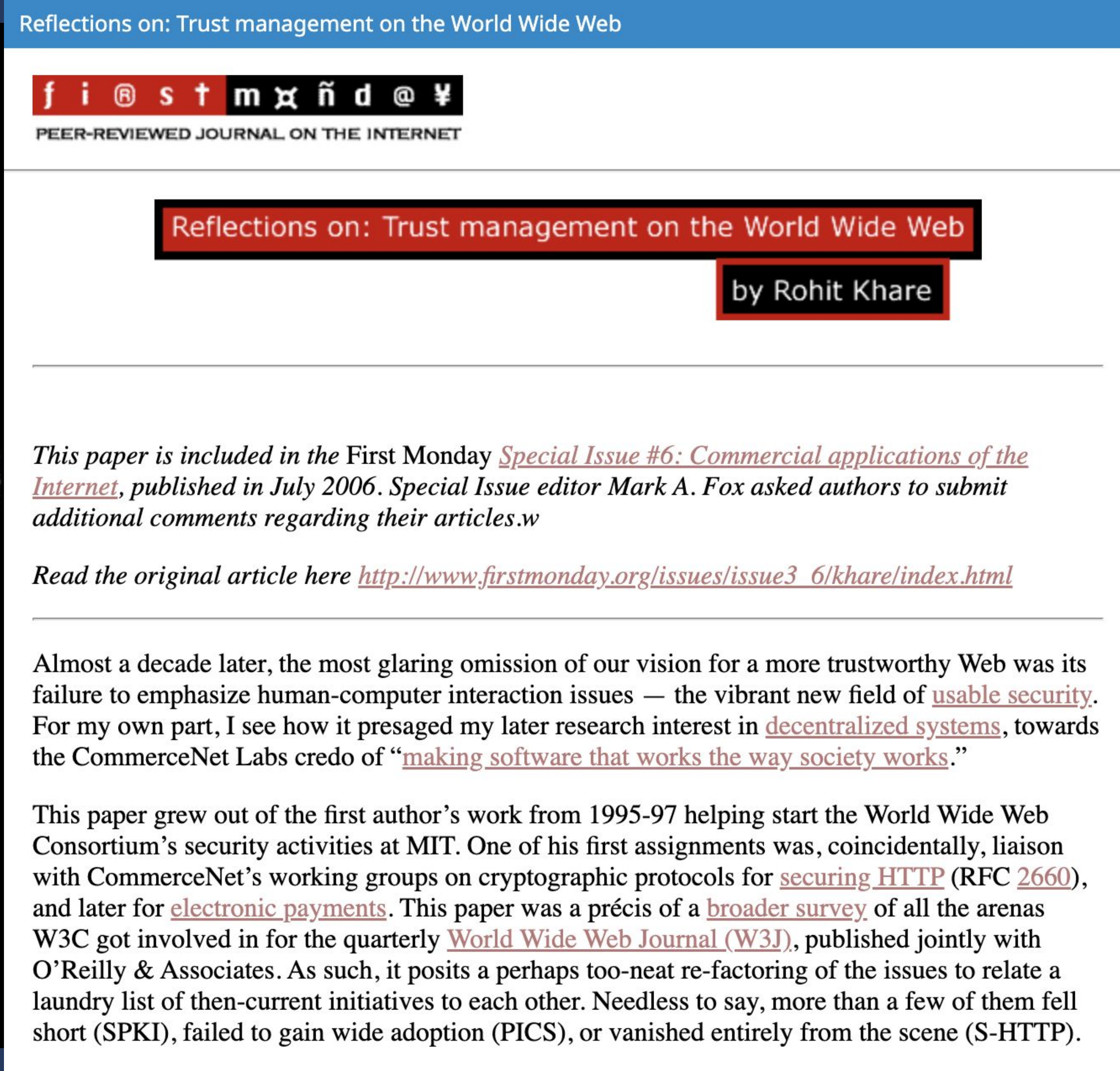
- Worked for ~~one~~ two Nobel Prize winners
- Hired by ~~one~~ two Turing Award winners
- On Wikipedia
- Translate API
- Pub/Sub GA
- Cloud IoT 1.0
- ML Engine
- IAM & Forseti
- Napa/Mesa

“Forrest Gump of the Web”

Middle-aged prodigy”

“Who?”




Reflections on: Trust management on the World Wide Web
f i ® s t m ¤ ñ d @ ¥
PEER-REVIEWED JOURNAL ON THE INTERNET

Reflections on: Trust management on the World Wide Web

by Rohit Khare

This paper is included in the First Monday *Special Issue #6: Commercial applications of the Internet, published in July 2006. Special Issue editor Mark A. Fox asked authors to submit additional comments regarding their articles.w*

Read the original article here http://www.firstmonday.org/issues/issue3_6/khare/index.html

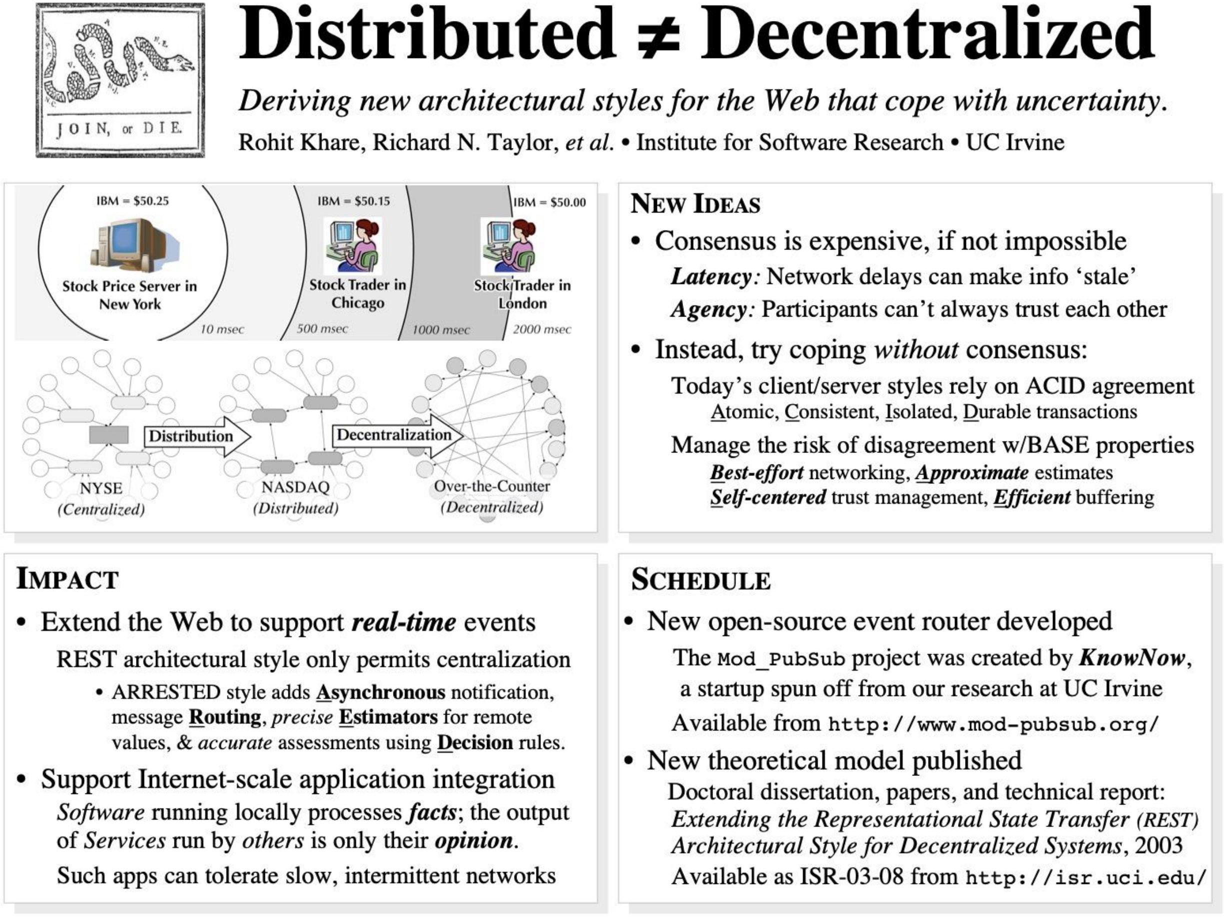
Almost a decade later, the most glaring omission of our vision for a more trustworthy Web was its failure to emphasize human-computer interaction issues — the vibrant new field of usable security. For my own part, I see how it presaged my later research interest in decentralized systems, towards the CommerceNet Labs credo of "making software that works the way society works."

This paper grew out of the first author's work from 1995-97 helping start the World Wide Web Consortium's security activities at MIT. One of his first assignments was, coincidentally, liaison with CommerceNet's working groups on cryptographic protocols for securing HTTP (RFC 2660), and later for electronic payments. This paper was a précis of a broader survey of all the arenas W3C got involved in for the quarterly World Wide Web Journal (W3J), published jointly with O'Reilly & Associates. As such, it posits a perhaps too-neat re-factoring of the issues to relate a laundry list of then-current initiatives to each other. Needless to say, more than a few of them fell short (SPKI), failed to gain wide adoption (PICS), or vanished entirely from the scene (S-HTTP).

Putting the Puzzle Pieces in Place:

Envisioning a Decent IAM…

(Excerpt from my 2003 graduate research)


JOIN, or DIE.
IBM = $50.25
Stock Price Server in New York
10 msec
IBM = $50.15
Stock Trader in Chicago
500 msec
1000 msec
IBM = $50.00
Stock Trader in London
2000 msec
Distribution
Decentralization
NYSE
(Centralized)
NASDAQ
(Distributed)
Over-the-Counter
(Decentralized)

Distributed ≠ Decentralized

Deriving new architectural styles for the Web that cope with uncertainty.

Rohit Khare, Richard N. Taylor, *et al.* • Institute for Software Research • UC Irvine

New Ideas

- Consensus is expensive, if not impossible

  Latency: Network delays can make info 'stale'

  Agency: Participants can't always trust each other

- Instead, try coping *without* consensus:

  Today's client/server styles rely on ACID agreement
  Atomic, Consistent, Isolated, Durable transactions

  Manage the risk of disagreement w/BASE properties
  Best-effort networking, ***Approximate*** estimates
  Self-centered trust management, ***Efficient*** buffering

Impact

- Extend the Web to support ***real-time*** events

  REST architectural style only permits centralization

  - ARRESTED style adds **Asynchronous** notification, message **Routing**, *precise* **Estimators** for remote values, & *accurate* assessments using **Decision** rules.

- Support Internet-scale application integration

  Software running locally processes ***facts***; the output of *Services* run by *others* is only their ***opinion***.

  Such apps can tolerate slow, intermittent networks

Schedule

- New open-source event router developed

  The `Mod_PubSub` project was created by ***KnowNow***, a startup spun off from our research at UC Irvine

  Available from `http://www.mod-pubsub.org/`

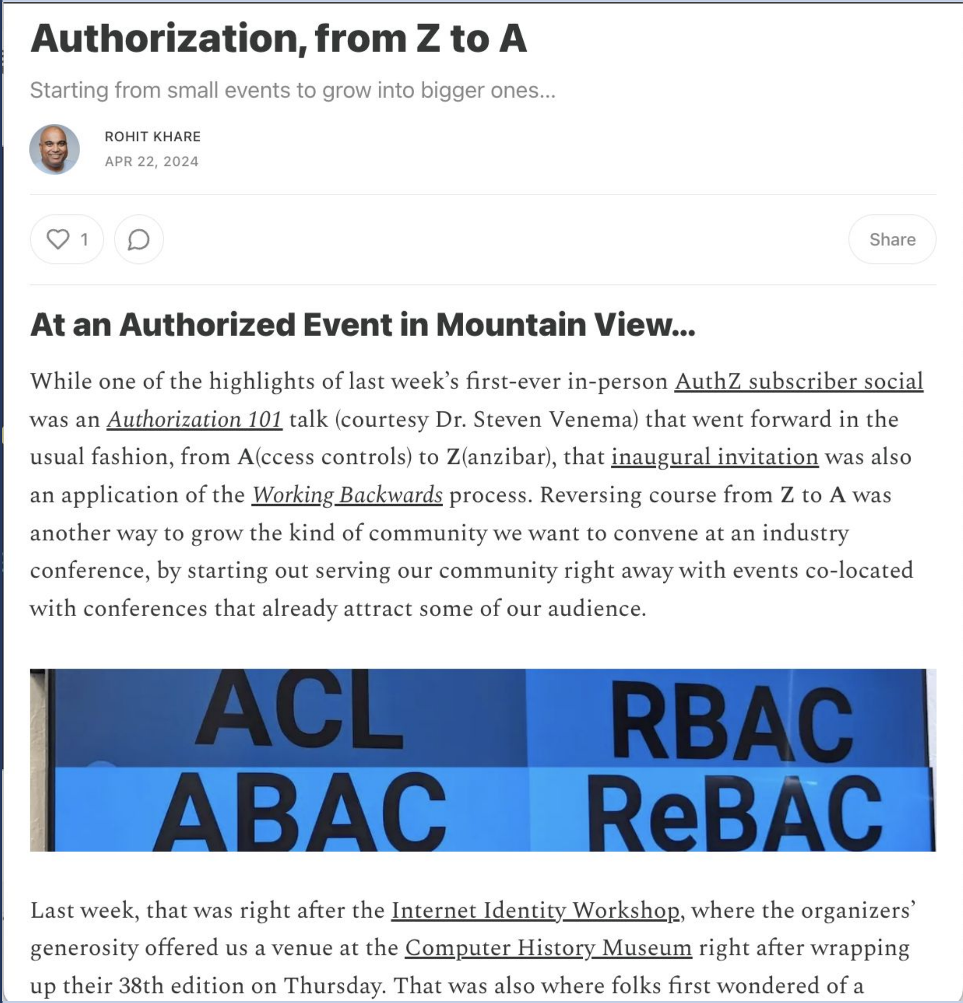
- New theoretical model published

  Doctoral dissertation, papers, and technical report:
  Extending the Representational State Transfer (REST) Architectural Style for Decentralized Systems, 2003
  Available as ISR-03-08 from `http://isr.uci.edu/`

2. What I've Learned

AuthZ Clipping Service

authz.substack.com



Authorization, from Z to A

Starting from small events to grow into bigger ones...

ROHIT KHARE
APR 22, 2024

At an Authorized Event in Mountain View...

While one of the highlights of last week's first-ever in-person AuthZ subscriber social was an *Authorization 101* talk (courtesy Dr. Steven Venema) that went forward in the usual fashion, from **A**(ccess controls) to **Z**(anzibar), that inaugural invitation was also an application of the *Working Backwards* process. Reversing course from **Z** to **A** was another way to grow the kind of community we want to convene at an industry conference, by starting out serving our community right away with events co-located with conferences that already attract some of our audience.

Last week, that was right after the Internet Identity Workshop, where the organizers' generosity offered us a venue at the Computer History Museum right after wrapping up their 38th edition on Thursday. That was also where folks first wondered of a

A...

... is for ***Authorization***, which spells out "*who can do what*"

— yet can't tell us "*everyone who could do something*"?

— nor whether there's "*someone who can do anything*"?

— or even explain "*everything someone could do*"?

B…

… is for ***“Bond,* James Bond,”** identified as an importer & exporter who wouldn’t scare me.

Parade

SUNDAY, SEPTEMBER 26, 2021 | PARADE.COM

THE BEST OF BOND

Who’s the deadliest 007, the creepiest villain, the most marvelous Moneypenny? Turn to page 10 to find out!

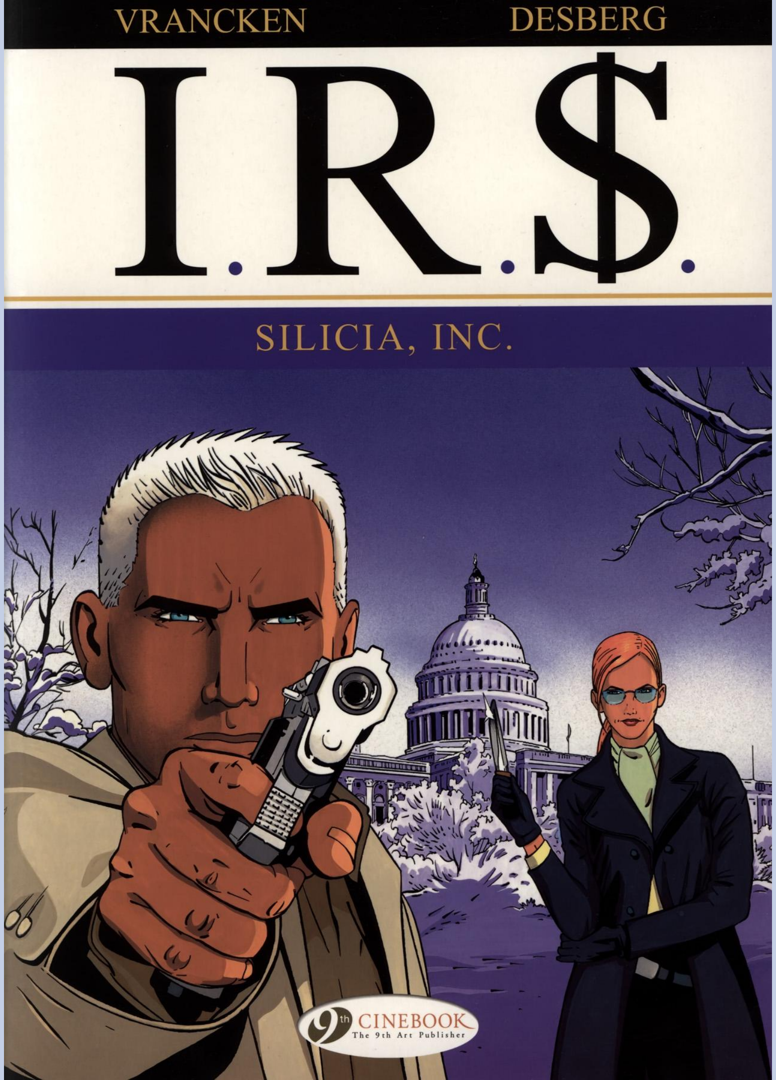
PLUS *Blue Bloods’* Tom Selleck, p. 2

TIMOTHY DALTON 1987–89
SEAN CONNERY 1962–71
DANIEL CRAIG 2006–21
PIERCE BROSNAN 1995–2002
GEORGE LAZENBY 1969
ROGER MOORE 1973–85

C…

… is for **Carry**, concealing his handguns and a *License to Kill* — run for the hills!

“Larry B. Max is an unusual specialist … Reading into tax-evasion and money-laundering rings the way a virtuoso pianist would read Mozart”



D…

… is for ***Deny***, the default path to security.

— If only there were NoBAC, there would be no breaches or attacks

… although our *developer velocity* would be severely set back!

— AuthZ architectures tradeoff privilege against productivity

E…

… is for ***Error***, a message that explains nothing at all.

— … since helpful hints for how to ask for access might aid an employee… or an enemy!

— Least Privilege so *strict* it wouldn’t even reveal what went wrong, nor how to ask for access

F…

… is for ***Formal Methods***, that use logic to replace *Intuition* with *Invariants*

— Enforced by *pre*-crime promises from Theorems turned into Laws, pure and stoic!

— Avoiding ACLs with Turing tar-pits is the critical insight of Amazon Cedar, AWS' approach to Verified Permissions and Access

— AWS' Zelkova automated reasoning dates back to 2018

G…

… is for **Git**, or
it didn't happen!

— for when an insider appears with access ill-gotten, where is the administrative paper trail tracking all of those approvals and inactions?

— *Security Policies are Software, Too*, harkening back to the origin of version control research in *Software Process are Software, Too*

H…

… is for the **Humans**, who command (and abandon) devices by the dozen

… and since each device and each computer has containers and calls cascading to yet other services on their behalf, that means chasing down more ***Non-**Human Identities*

– NHI standards like SPIFFE and SPIRE may be based on digital signatures from PKIs, but it’s a WIMSE-ical adventure to map between people and processes…

Workload Identity in Multi System Environments

I...

... is for **Intelligence,**
Artificial at best,
Superficial at worst!

— AI appears everywhere today
... except in actual productivity?

— Agents need to incorporate *only* the information users are authorized to access, *before* answering or acting!

— Forgetting is the hardest part

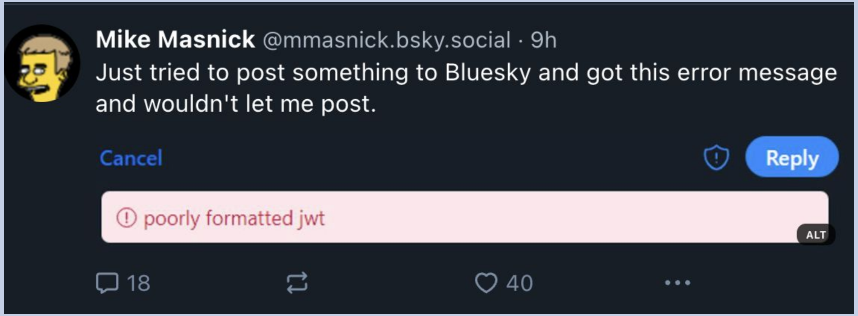
J…

… is for **Jot**, spelled ‘JWT’ and as short as a tittle, yet can delegate access like a drunk after a tipple!

— Digitally-signed Web tokens in AuthN/Z protocols differ from PKI for digital signatures

— Much more secure than cookies, yet still can cause confused deputies if careless

— They are close cousins to pre-signed URLs for ‘temporary’ storage access: Azure AI accident



Mike Masnick @mmasnick.bsky.social · 9h

Just tried to post something to Bluesky and got this error message and wouldn't let me post.

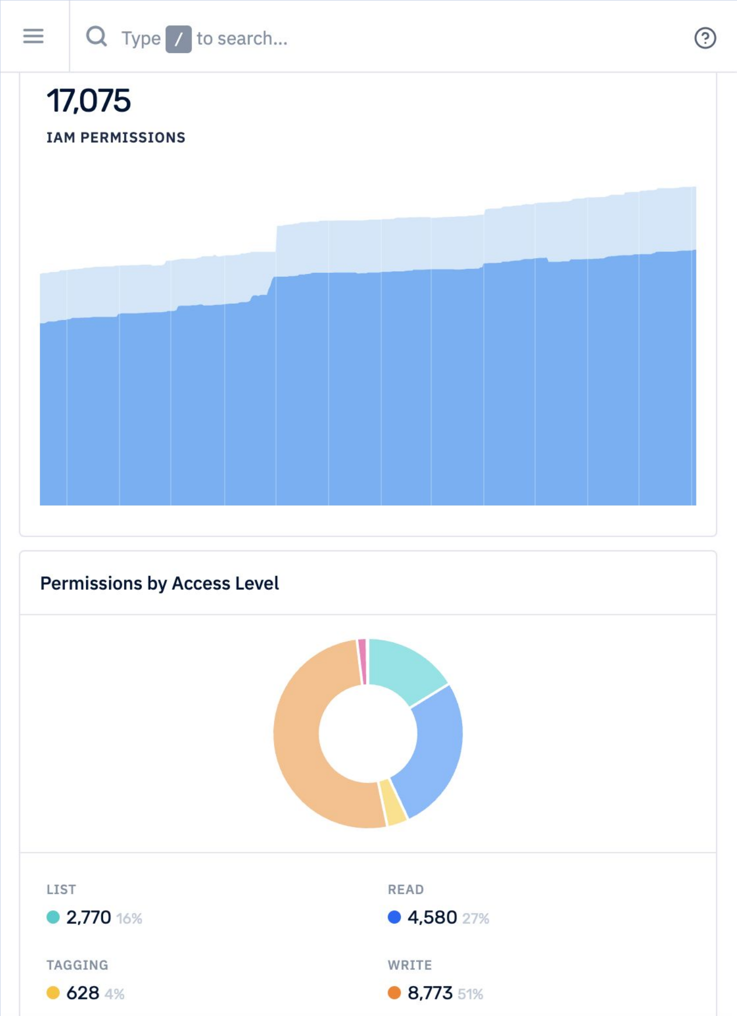
Cancel Reply

ⓘ poorly formatted jwt

K...

... is for **Knowledge**, of the true power of permissions finer-grained than flour, yet obsoleted within hours

permissions.cloud



17,075

IAM PERMISSIONS

Permissions by Access Level

LIST
● 2,770 16%

READ
● 4,580 27%

TAGGING
● 628 4%

WRITE
● 8,773 51%

L…

… is for **Language**, with which we are cursed by the dozen, lacking any Rosetta Stones for cousins

— From XACML to ALFA, with IDQL, and OPA with Rego and CEL, not to mention Kogito, Drools, and clones of Zanzibar…

— Boundaries are blurred between policy engines and “their” own policy languages (e.g. Topaz is multi-language)

M…

… is for **Mandatory**, which some access controls aspire to be

— Locking out meddling kids from configuring controls can be necessary, mostly in the military

— Today, almost all ACLs would be considered ‘Discretionary’

… Since careful configuration can still enforce segregation of duties

North Star – Open Policy Agent – Policy – Questions

N… is for the **North Star** that guides us on the path to *least-privilege* and helps shift-left and shift-closer every night.

O… is for **Open Policy Agent,** an enforcer-in-a-box who rules the land of Kubernetes, but emerged from the tar-pits of Turing.

P… is for **Policy,** at the Point of Enforcement, or Administration, or Information, or even Decisions

Q… is for **Questions** that cannot yet be asked or answered, like *“How many interns could copy passports from passengers?”*

ReBAC – Superusers – Tags – Units

R… is for **ReBAC,** from the promised island of *Zanzibar*, where you ask not what you can do on your own, but how each action associates one to another

S… is for **Superusers,** cosplay heroes with Kleene stars on their shields who save the day, but by breaking glass and run roughshod over the path to least privilege

T… is for a **Tag,** which is "It" for ABAC, begging the question: "*Who will watch the taggers? Banksy knows*, *and arranges accordingly…*"

U… is for a **Unit,** *Organizational* or otherwise, aggregating and inheriting policies leaving it clear as mud which are going to apply

Verified Permissions – Wiki – X – Why?

V… is for **Verified Permissions,** the most conspicuously esoteric and advanced computer science that Amazon's ever sold

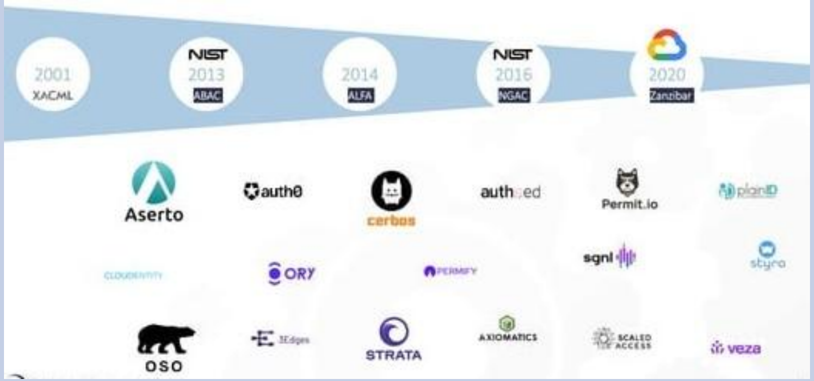
W… is for **Wiki,** where all can contribute, quickly and equally, to all the laws and guardrails we are always adjudicating and altering

X… is for **X**, that variable worth solving for, to *prove who* could do *what*, to *which* things, and *when*?

Y… is for **Why?,** the most plaintive call of all, after breaches are reconstructed from forensic files that failed to log what went wrong, while it was happening.

Z…

… is for ***Zen***, a state with checks all in balance

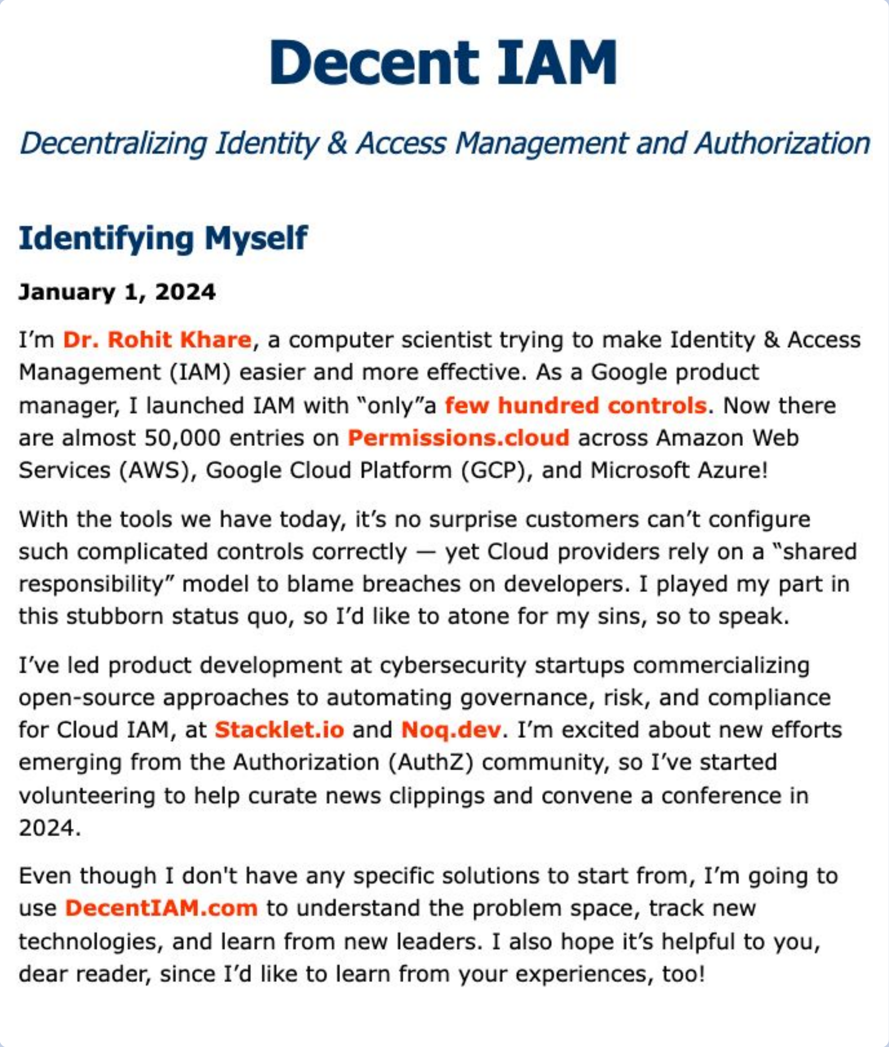


— once *AuthZEN* allows all access controls to be written in any policy language, integrating all of the pieces, and all at peace.

3. What I'm Exploring

Decentralizing IAM & AuthZ

DecentIAM.com



Decent IAM

Decentralizing Identity & Access Management and Authorization

Identifying Myself

January 1, 2024

I'm **Dr. Rohit Khare**, a computer scientist trying to make Identity & Access Management (IAM) easier and more effective. As a Google product manager, I launched IAM with "only"a **few hundred controls**. Now there are almost 50,000 entries on **Permissions.cloud** across Amazon Web Services (AWS), Google Cloud Platform (GCP), and Microsoft Azure!

With the tools we have today, it's no surprise customers can't configure such complicated controls correctly — yet Cloud providers rely on a "shared responsibility" model to blame breaches on developers. I played my part in this stubborn status quo, so I'd like to atone for my sins, so to speak.

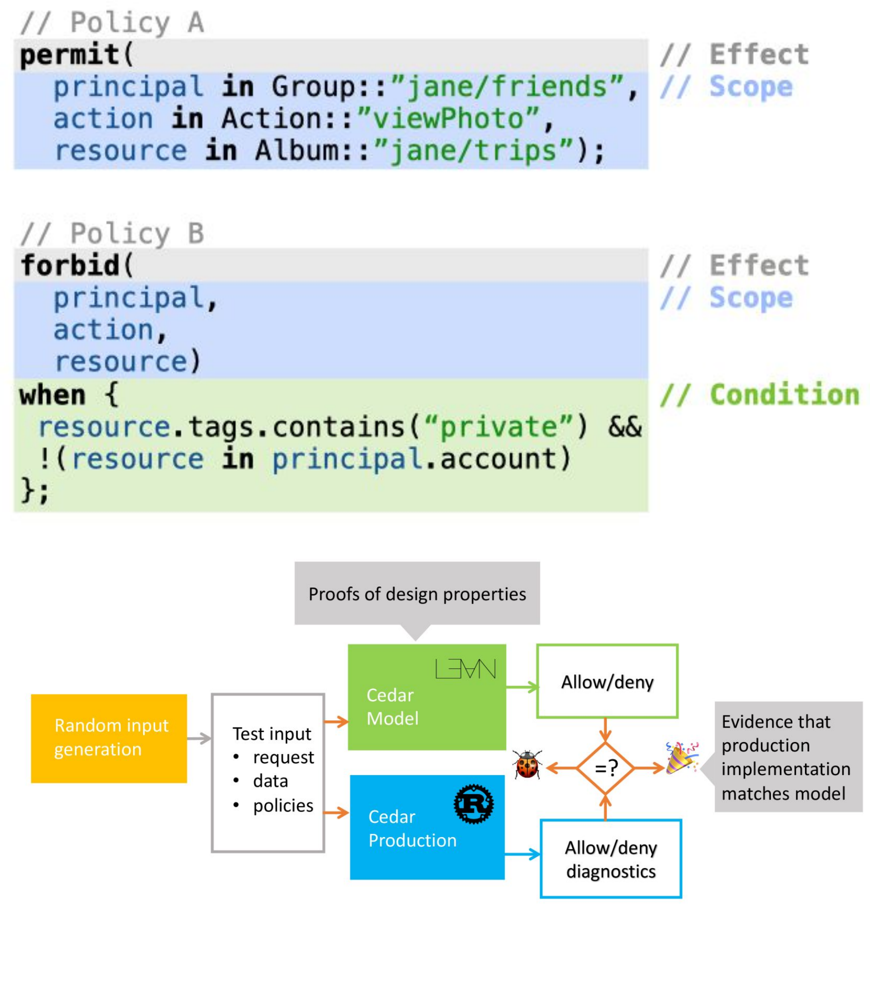
I've led product development at cybersecurity startups commercializing open-source approaches to automating governance, risk, and compliance for Cloud IAM, at **Stacklet.io** and **Noq.dev**. I'm excited about new efforts emerging from the Authorization (AuthZ) community, so I've started volunteering to help curate news clippings and convene a conference in 2024.

Even though I don't have any specific solutions to start from, I'm going to use **DecentIAM.com** to understand the problem space, track new technologies, and learn from new leaders. I also hope it's helpful to you, dear reader, since I'd like to learn from your experiences, too!

“Proven Policies” with Cedar

Access Control Lists, Permissions, Entitlements, Capabilities — all these alternatives are appealing

All can become analytically intractable if expressed using conditions, branches, or regular expressions, that aren’t *invertible*



```
// Policy A
permit(                                            // Effect
  principal in Group::”jane/friends”,              // Scope
  action in Action::”viewPhoto”,
  resource in Album::”jane/trips”);

// Policy B
forbid(                                            // Effect
  principal,                                       // Scope
  action,
  resource)
when {                                             // Condition
 resource.tags.contains(“private”) &&
 !(resource in principal.account)
};
```

Proofs of design properties

Random input generation

Test input
- request
- data
- policies

Cedar Model

Cedar Production

Allow/deny

Allow/deny diagnostics

=?

Evidence that production implementation matches model

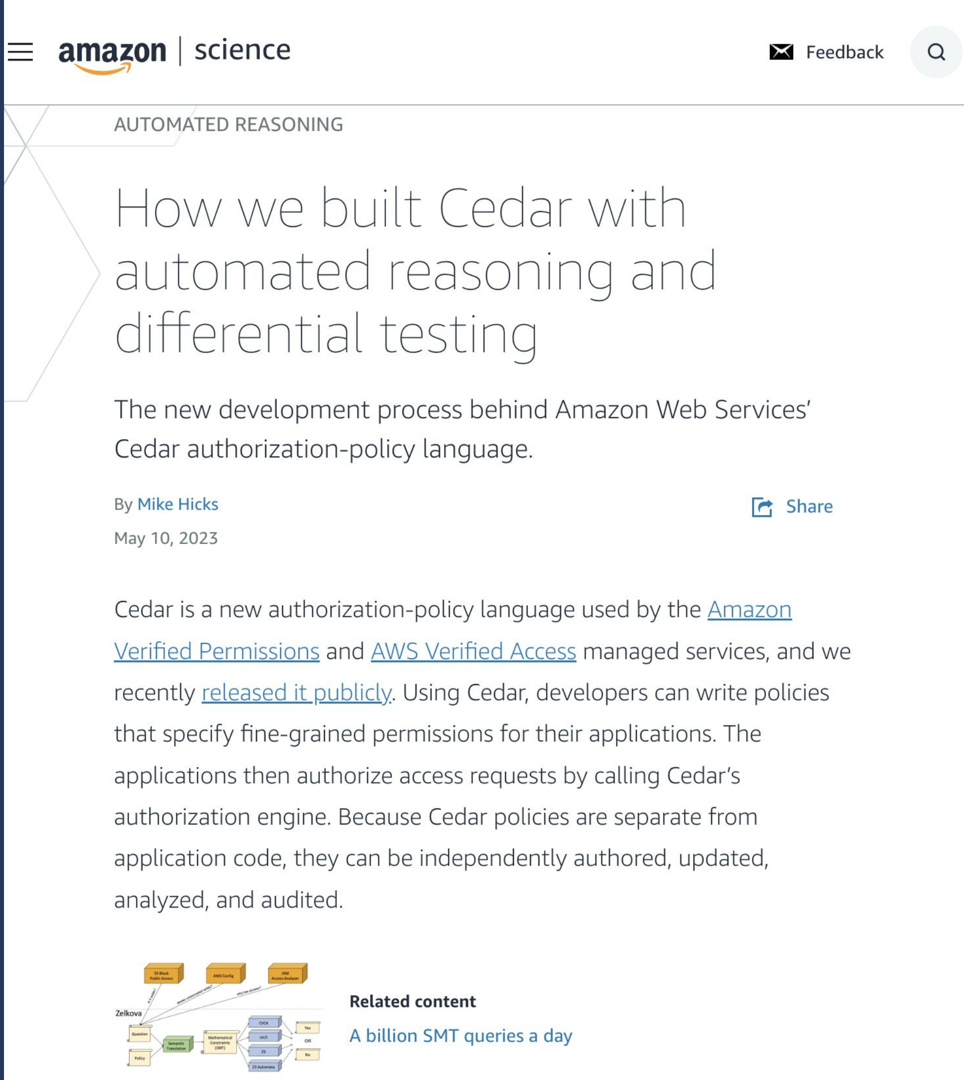
"*Reasoning*" with *Solvers*

The open-source playground at *CedarPolicy.com* has one unique implementation, written in *Dafny*

That means there are formal semantics to *grant* or *deny*.

Of course, verification of an abstraction does not mean it's actually a secure *implementation*!

Rust is fuzz-tested 10^8 times/day



AUTOMATED REASONING

How we built Cedar with automated reasoning and differential testing

The new development process behind Amazon Web Services' Cedar authorization-policy language.

By Mike Hicks

Share

May 10, 2023

Cedar is a new authorization-policy language used by the Amazon Verified Permissions and AWS Verified Access managed services, and we recently released it publicly. Using Cedar, developers can write policies that specify fine-grained permissions for their applications. The applications then authorize access requests by calling Cedar's authorization engine. Because Cedar policies are separate from application code, they can be independently authored, updated, analyzed, and audited.

Related content

A billion SMT queries a day

“Plain English” with *Chatbots*

Turning natural language into code risks creating hallucinations

… but bad policies are even riskier

So instead of using language models to generate ACLs, imagine a co-pilot for creating constraints

And instead of answering security audits, edit the answers to suit…

“Interns can’t access passports”

→ *What’s an intern?*

“Interns@ can’t read passports”

→ *What about test passengers?*

“Interns@ can’t query any tables with passport numbers from production systems”

→ *What if managers approve?*
→ *What about support cases?*

“Interns@ can only query any tables with passport numbers in production when manager approves & passenger consents”

“Version Control” with *Iambic*

At Noq, we open-sourced a new language for “IAM, *but in code*”

Deployed faster than Terraform

Even temporary escalated access appears in auditable commit logs

Simulate impact of policy changes by replaying past CloudTrail logs

GitHub.com/noqdev/iambic

```yaml
template_type: NOQ::Okta::Group
idp_name: main
iambic_managed: enforced
properties:
  name: engineering_interns
  description: Engineering Interns
  members:
    - username: intern1@foo.com
      expires_at: 2024-09-01
         # Interns last day
    - username: intern2@foo.com
      expires_at: 2024-11-01
```

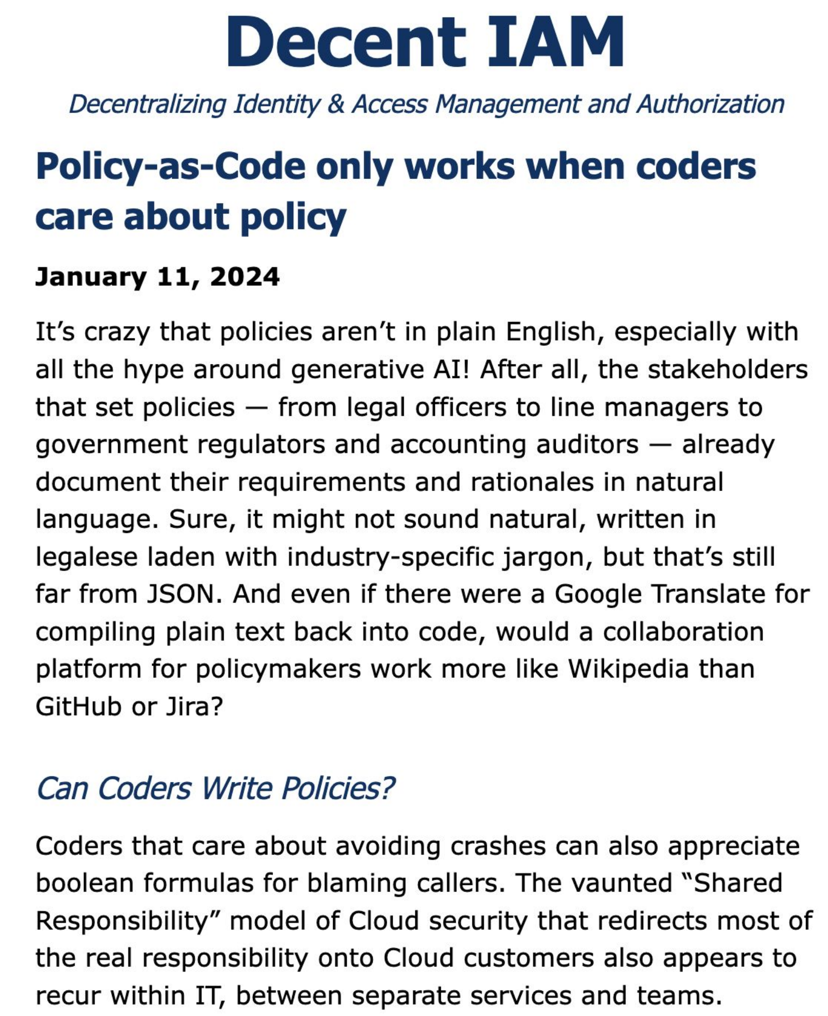
“Collaborate” with *Policypedia*

The original sin of Policy-as-Code is *coders don’t care about policy*

An open-source repository of controls is only legible to Devs

Other stakeholders deserve docs clear enough to edit as peers

Every policy has an exception, so every page deserves debates, too.



Decent IAM

Decentralizing Identity & Access Management and Authorization

Policy-as-Code only works when coders care about policy

January 11, 2024

It’s crazy that policies aren’t in plain English, especially with all the hype around generative AI! After all, the stakeholders that set policies — from legal officers to line managers to government regulators and accounting auditors — already document their requirements and rationales in natural language. Sure, it might not sound natural, written in legalese laden with industry-specific jargon, but that’s still far from JSON. And even if there were a Google Translate for compiling plain text back into code, would a collaboration platform for policymakers work more like Wikipedia than GitHub or Jira?

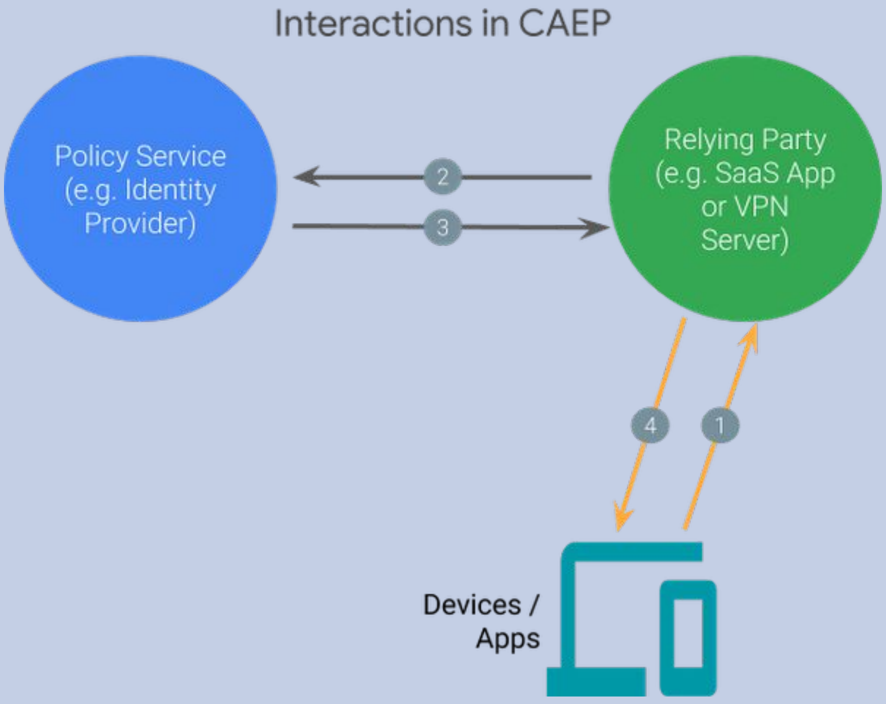
Can Coders Write Policies?

Coders that care about avoiding crashes can also appreciate boolean formulas for blaming callers. The vaunted “Shared Responsibility” model of Cloud security that redirects most of the real responsibility onto Cloud customers also appears to recur within IT, between separate services and teams.

"Elastic AuthN" with *CAEP*

For AuthN, Continuous Access Evaluation Protocol (CAEP) addresses the dimension of *time:*

- When an employee travels abroad; a key leaks; or taking a high-risk actions…
- … then ask the user to tap their token; switch to a corporate device; or escalate for manager approval

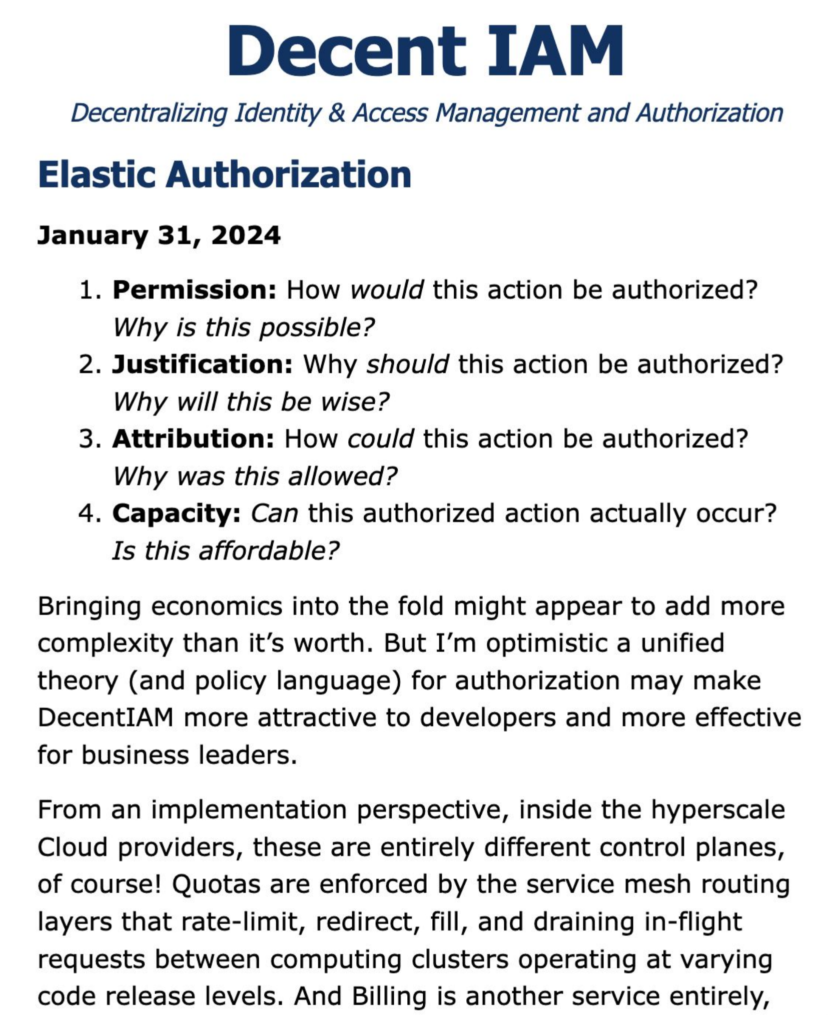

Interactions in CAEP
Policy Service (e.g. Identity Provider)
Relying Party (e.g. SaaS App or VPN Server)
2
3
4
1
Devices / Apps

caep.dev

"Elastic AuthZ" as Economics

For AuthZ, there's frontier beyond grant vs. deny: *quotas & budgets*:

- Do enough resources exist?
- Can you afford to pay for it?
- Have you tried it too often?
- Does it re-identify people?
- Have users consented to it?



Decent IAM

Decentralizing Identity & Access Management and Authorization

Elastic Authorization

January 31, 2024

1. **Permission:** How *would* this action be authorized? *Why is this possible?*
2. **Justification:** Why *should* this action be authorized? *Why will this be wise?*
3. **Attribution:** How *could* this action be authorized? *Why was this allowed?*
4. **Capacity:** *Can* this authorized action actually occur? *Is this affordable?*

Bringing economics into the fold might appear to add more complexity than it's worth. But I'm optimistic a unified theory (and policy language) for authorization may make DecentIAM more attractive to developers and more effective for business leaders.

From an implementation perspective, inside the hyperscale Cloud providers, these are entirely different control planes, of course! Quotas are enforced by the service mesh routing layers that rate-limit, redirect, fill, and draining in-flight requests between computing clusters operating at varying code release levels. And Billing is another service entirely,

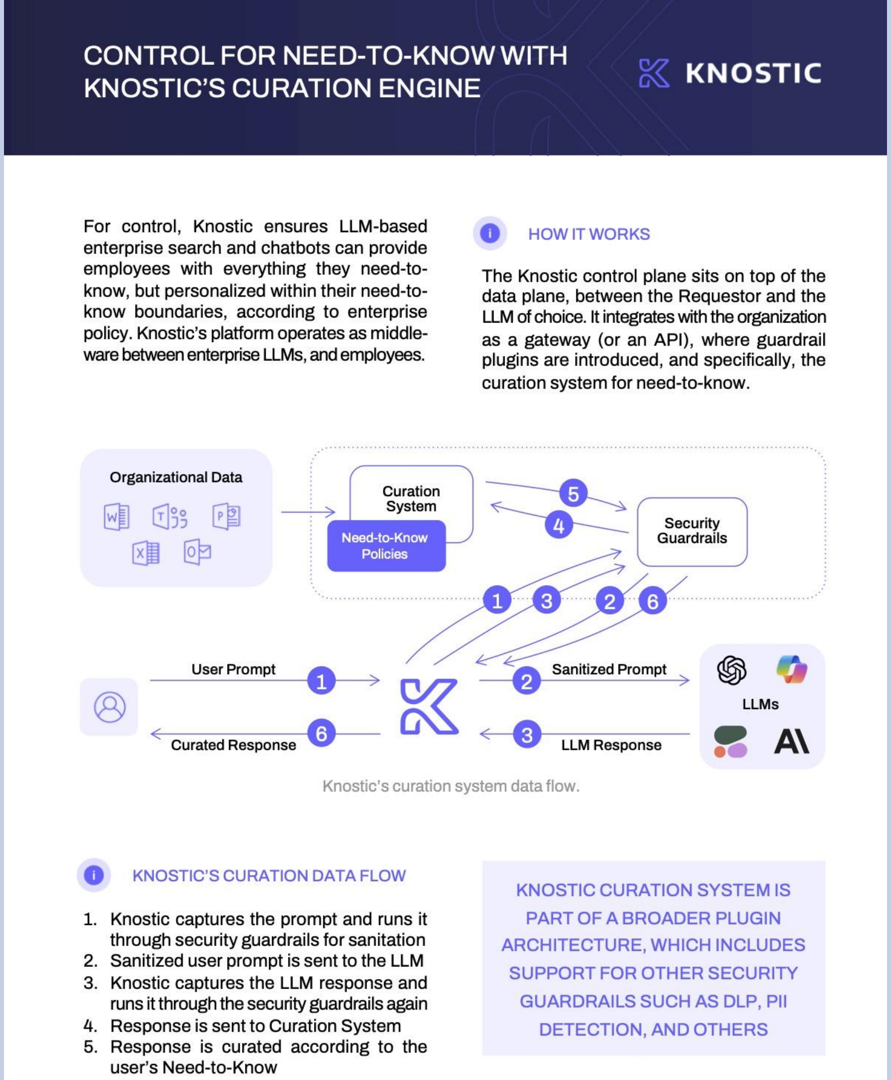
“Decentralization” for *Agents*

The “killer app” for a Decent IAM is search, or “permissioned RAG”

Importing information requires importing ACLs along with it

Enforcing ACLs without waiting for original systems to respond requires portable emulation

“*Google Translate*” for guardrails?



CONTROL FOR NEED-TO-KNOW WITH KNOSTIC’S CURATION ENGINE

KNOSTIC

For control, Knostic ensures LLM-based enterprise search and chatbots can provide employees with everything they need-to-know, but personalized within their need-to-know boundaries, according to enterprise policy. Knostic’s platform operates as middleware between enterprise LLMs, and employees.

HOW IT WORKS

The Knostic control plane sits on top of the data plane, between the Requestor and the LLM of choice. It integrates with the organization as a gateway (or an API), where guardrail plugins are introduced, and specifically, the curation system for need-to-know.

Knostic’s curation system data flow.

KNOSTIC’S CURATION DATA FLOW

1. Knostic captures the prompt and runs it through security guardrails for sanitation
2. Sanitized user prompt is sent to the LLM
3. Knostic captures the LLM response and runs it through the security guardrails again
4. Response is sent to Curation System
5. Response is curated according to the user’s Need-to-Know

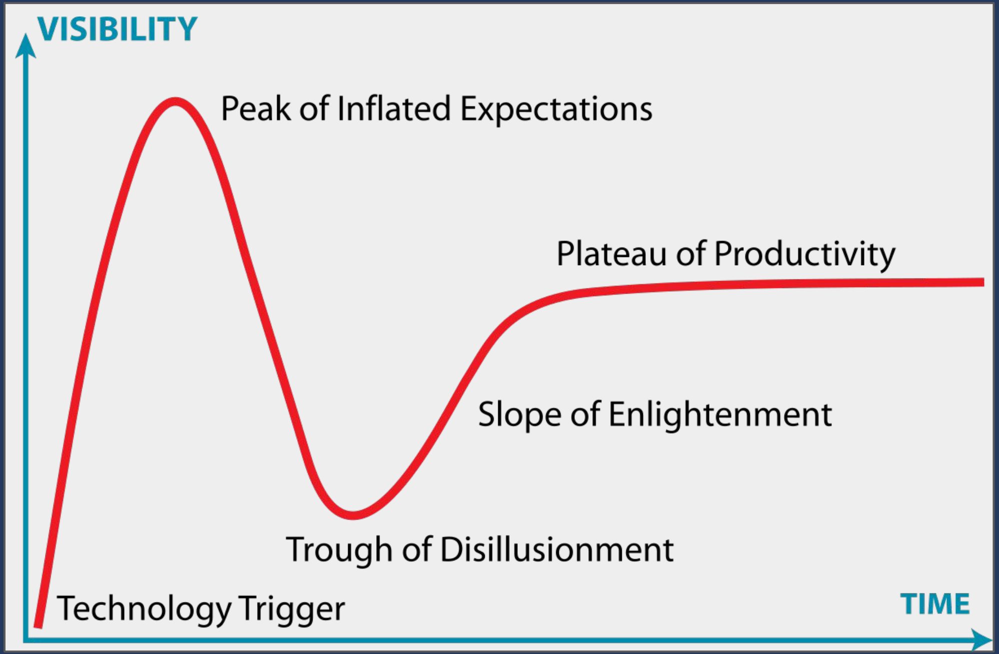
KNOSTIC CURATION SYSTEM IS PART OF A BROADER PLUGIN ARCHITECTURE, WHICH INCLUDES SUPPORT FOR OTHER SECURITY GUARDRAILS SUCH AS DLP, PII DETECTION, AND OTHERS

Drawing Conclusions for the Future of AuthZ

External AuthN is “easier” to sell, since all apps have users

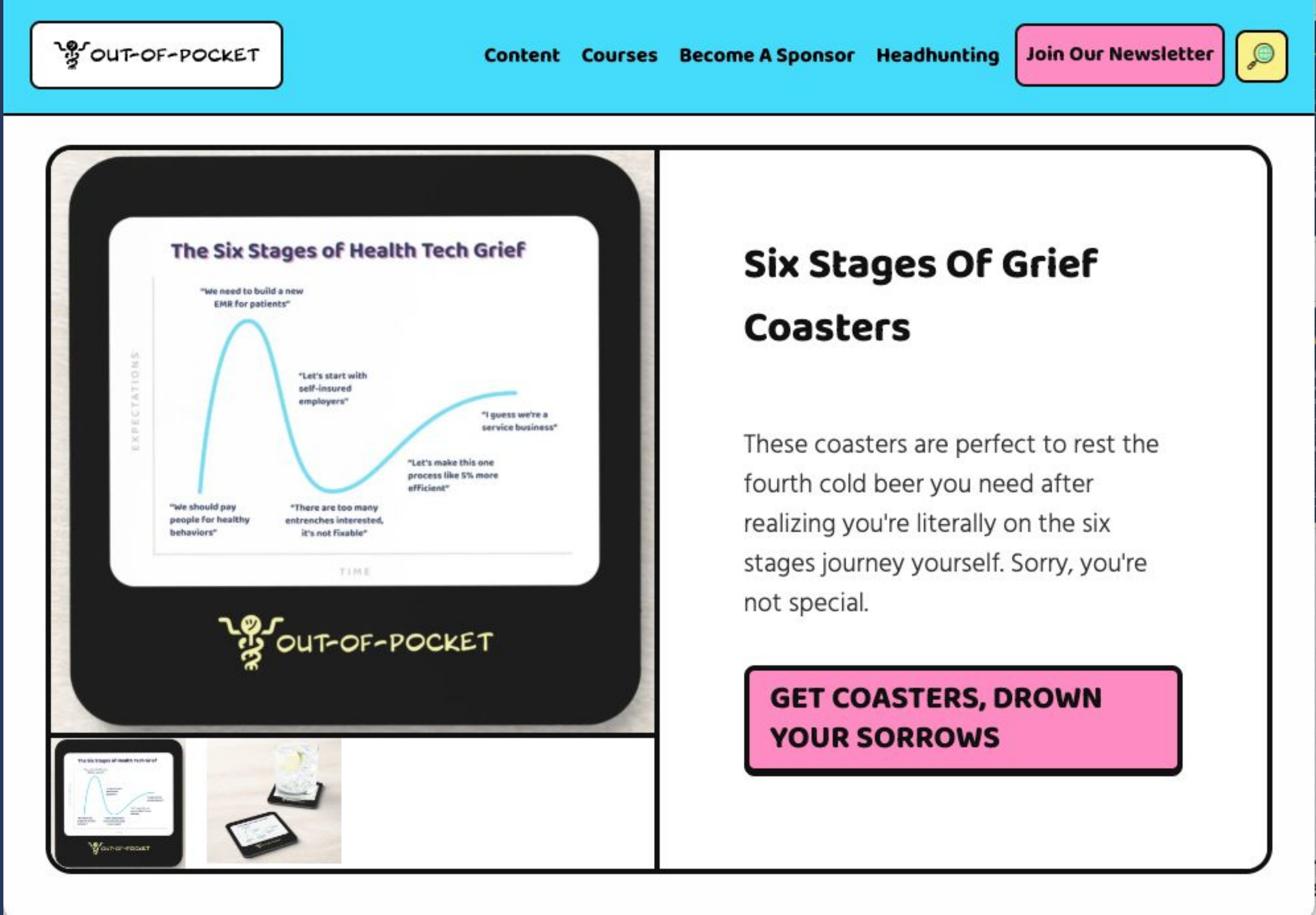
External AuthZ is “harder” to sell, since problems with permissions only appear later, accumulating over time

Converting teams one app at a time may take longer to adopt than integrating ACLs



My Role Model:

"Health care, but funny!"



OUT-OF-POCKET
Content
Courses
Become A Sponsor
Headhunting
Join Our Newsletter
The Six Stages of Health Tech Grief
"We need to build a new EMR for patients"
"Let's start with self-insured employers"
"I guess we're a service business"
"Let's make this one process like 5% more efficient"
"We should pay people for healthy behaviors"
"There are too many entrenched interested, it's not fixable"
EXPECTATIONS
TIME
OUT-OF-POCKET

Six Stages Of Grief Coasters

These coasters are perfect to rest the fourth cold beer you need after realizing you're literally on the six stages journey yourself. Sorry, you're not special.

GET COASTERS, DROWN YOUR SORROWS