



id: SI_005

model: anthropic/claude-sonnet-4.6

seeds:

04_risiko: sårbarhet

03_relasjon: makt

05_handling: utforsk

01_perspektiv: operasjonell

02_tidhorisont: kortsiktig

Instans SI_005

Seeds: operasjonell | kortsiktig | makt | sårbarhet | utforsk

Maktforskyvning er den egentlige risikoen

Mainstream-diskusjonen handler om teknisk kontroll. Men den operasjonelle realiteten er en annen: **agentisk KI omfordeler makt** – fra mennesker til systemer, fra revisjonslogikk til prosesslogikk. Det er dette som er den kortsiktige sårbarhetsvektoren dere undervurderer.

Med distribuert logging, svak IAM og ingen policy-motor har dere ikke en kontrollert autonomi-pilot. Dere har et **makteksperiment uten sikkerhetsnett**.

Operasjonelle blindsoner som slår til innen 6-9 måneder

1. Ephemeral credentials uten revokasjon-pipeline er ikke bare en IAM-svakhet – det er en maktlekkasje. En agent med utdaterte, men gyldige tokens kan operere utenfor tiltenkt kontekst i timevis. Implementer TTL < 15 minutter og tving rotasjon via Vault eller AWS IAM Roles Anywhere fra dag én.

2. Legacy-API-er uten idempotens er den farligste sårbarhetssonen kortsiktig. Én agent-retry mot et fagsystem uten idempotency-key kan doble en utbetaling, slette en post eller trigge en varslingskaskade. Krev eksplisitt idempotency-dokumentasjon for alle API-er agenten får nå – og blokkér de udokumenterte.

3. Latent coupling i eldre monolitter aktiveres uforutsigbart. En agent som trigger hendelse X i system A kan via databasetriggere aktivere prosess Y i system B som ingen visste var koblet. Dette er emergent behavior som ikke fanges i unit-testing. Utforsk produksjonstrafikk med chaos-injeksjon før agenten rulles ut.



Makt uten observability er ukontrollert makt

Uten konsolidert event-sourcing har dere ingen operasjonell evne til å svare på: Hva bestemte agenten, og hvorfor, kl. 03:47? Minimal løsning nå: Structured logging med korrelasjon-ID per agent-kjøring, sendt til én sentral sink (f.eks. OpenSearch). Ikke perfekt – men operasjonelt tilstrekkelig.

Konkrete neste steg

1. **Uke 1-2:** Kartlegg alle API-er agenten skal nå. Klassifiser etter idempotens-status. Blokkér ikke-idempotente endepunkter fra pilot-scope.
2. **Uke 3-4:** Deploy Vault for ephemeral credentials med 15-minutters TTL. Ingen agent får langlivede service-account-nøkler.
3. **Måned 2:** Gjennomfør én strukturert “blast radius”-øvelse: simuler at agenten feiler midt i en transaksjonskjede og mål faktisk gjenopprettingstid.

Utforsk det ubehagelige: hva skjer operasjonelt når agenten har rett, men systemet er feil?