Towards a Scalable and Pluggable Data Policy System
Will Wang & Rachel Li, University of California, Berkeley

Motivation

- Recent regulations like GDPR and CCPA make it important that data is correctly stored and used
- Data is difficult to manage in large, fast growing distributed microservice applications
- RBAC for data in a distributed system is not enough, because data flows from service to service and RBAC is only efficiently enforced at the DB level

Goals

- Build a system where every piece of data is tagged with a label
- Services have externally managed privilege annotations dictating what kinds of data they can access
- Services that do not have the right privileges cannot access privileged data. This enforces least privilege and makes sure that data is not inadvertently used in dangerous ways

Constraints:

- Build with currently available, industry standard distributed application tools
- Require little to no modifications to the preexisting application - all functionality handled by the infrastructure
- Can trust application to have correct mechanism to forward labels, but cannot trust application to perform access control

System Design

Our system is built on top of the container orchestration Kubernetes, the service mesh Istio, and the reverse proxy Envoy.

- Use applications that implement distributed tracing, since they already have the mechanism to forward a request-id header to downstream
- An HTTP header (x-data) contains label information and is passed between services by the Envoy reverse proxy
- If a service with privileged x-data arrives at an unprivileged application, it is blocked by Envoy before the application receives that data

Experimental

We experimented by adapting modifying a pre-existing example microservice application built by Istio. We also ran some microbenchmarks.

- 1000 requests at 50 concurrency through a single HTTP proxy
- +165% over no proxy
- +26.8% over default proxy

99th Percentile Request Latency

- No Service Mesh
- Istio + Envoy
- Istio + Envoy with Filter

1000 requests at 50 concurrency on the Bookinfo app under different set-ups.

- +2% over no mesh
- +0.4% over default mesh

Latency through Single Proxy

- No Proxy
- Envoy Proxy
- Envoy Proxy with Filter

7.345 ms 3.3566 ms 3.3716 ms