Oblivious HTTP Deployment Experiences

Tommy Pauly
OHAI
July 2024, Vancouver

Agenda

Use cases

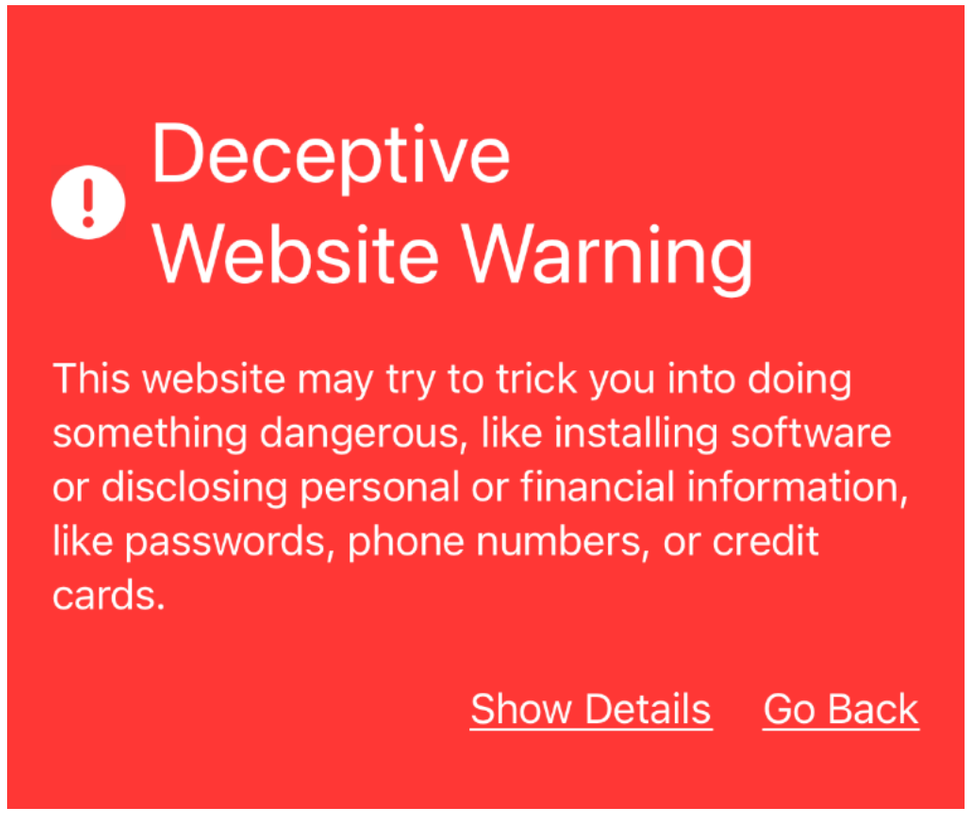
Configuration model

Use Case 1: Safe Browsing

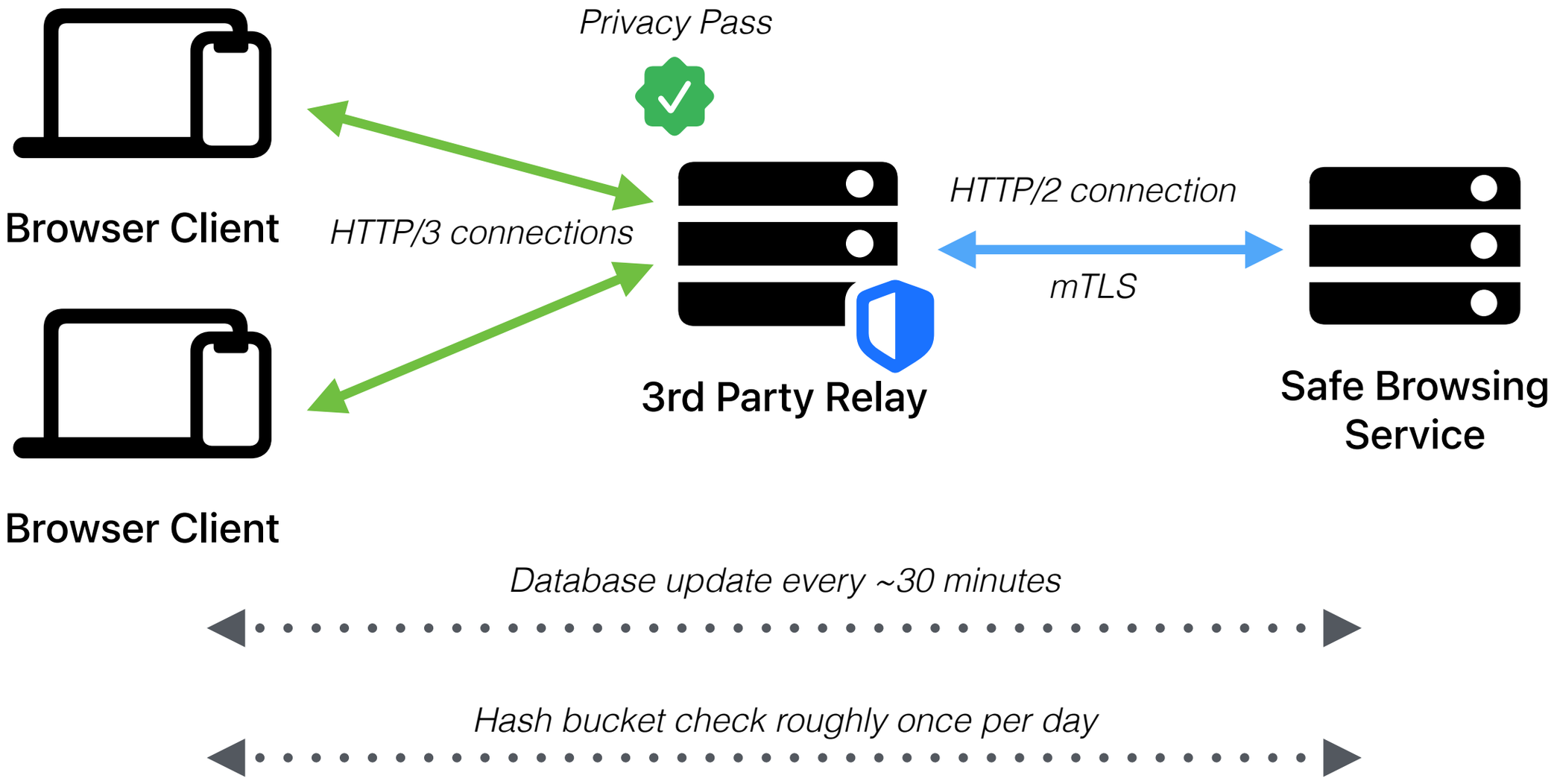
Periodic background fetches of a list of truncated URL hashes to detect phishing sites, etc.

Requests for complete hashes when a potential match is found

Needs privacy to avoid servers tracking all user IP addresses and potentially correlating activity



Safe Browsing



Safe Browsing

Normally a "cold start" connection

Predominantly background, latency is not blocking users

P50 times globally are around 350ms for database download

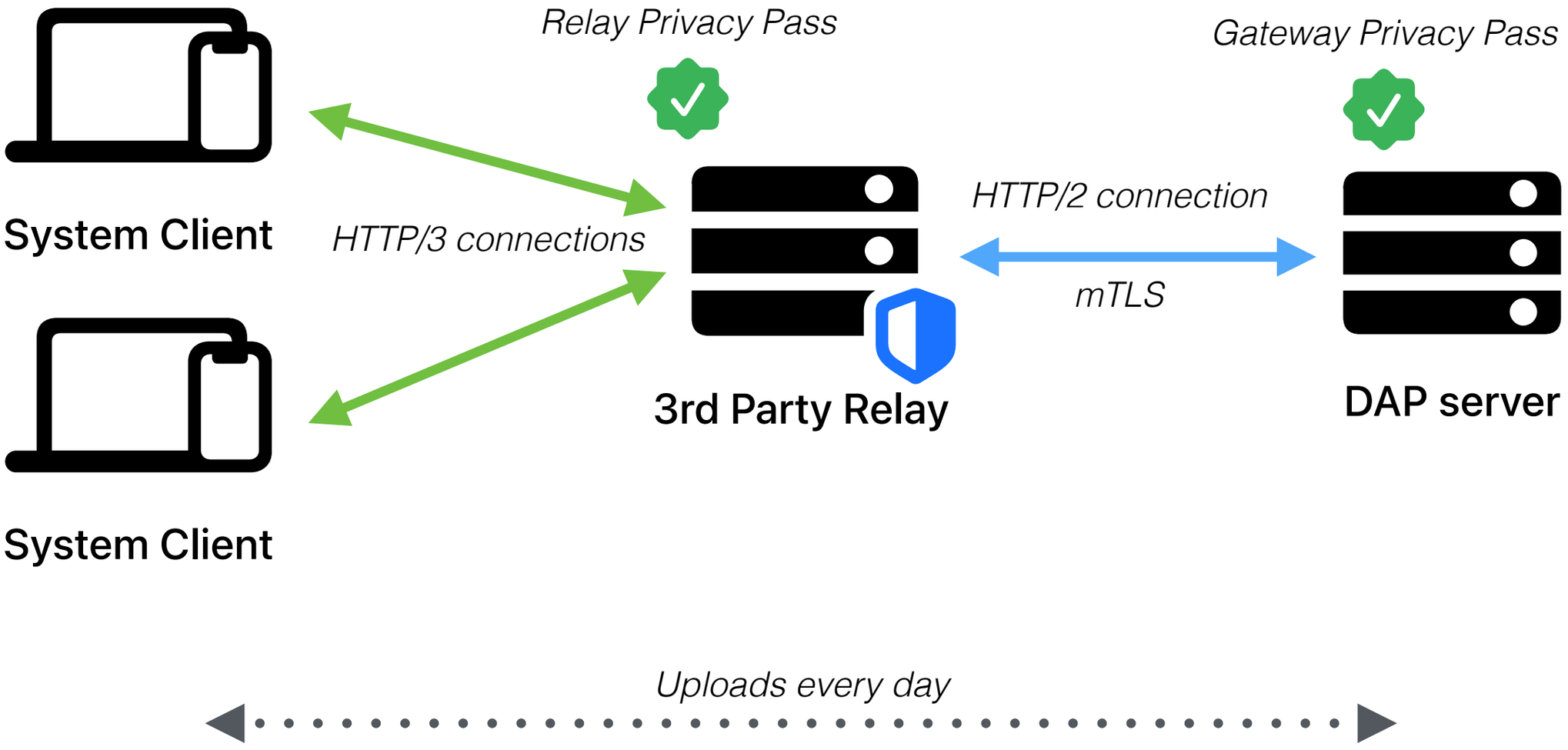
Use Case 2: Metrics Upload

DAP metrics upload over OHTTP

Responses are not used by client

- This is a case that could be optimized by having the relay ACK the message and forward it along ("unreliable")

Metrics Upload



Uploads every day

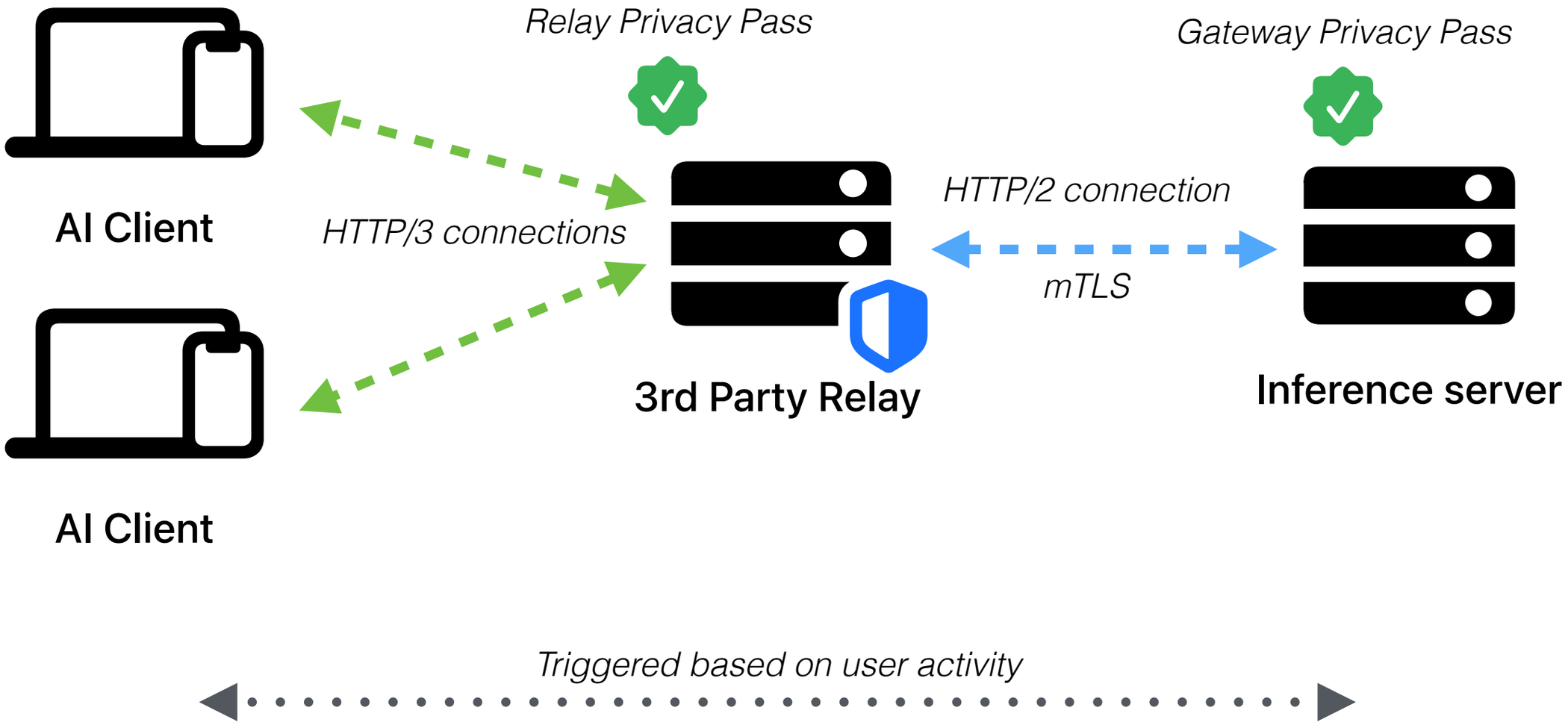
Use Case 3: Model Inference

Chunked OHTTP is being used for Private Cloud Compute and related AI features

https://security.apple.com/blog/private-cloud-compute/

Responses can be parsed in segments, and take a varying amount of time to generate

Model Inference



Model Inference

New aspects used with Chunked OHTTP

- Indeterminate length BHTTP messages
- BHTTP trailers used to indicate errors
- *These weren't exercised before and needed cleanup in client implementation*

Chunked OHTTP also exposed common CDN buffering behavior with POST

Deployment Model

Clients fetch a configuration bag every 12-24 hours

OHTTP rules include:

- Gateway key configuration
- Relay resource URLs
- Supported target resources
- Mode (non-chunked/chunked)

Some of these are included in a transparency log to reduce targeting ability

Questions?