The DECOUPLING Principle

Barath Raghavan
USC / INVISV

Thanks to: Paul Schmitt (INVISV), Jana Iyengar (Fastly),
Chris Wood (Cloudflare), Tommy Pauly (Apple)

In a Nutshell

For Internet privacy, decouple who you are from what you do

➔ Old idea (dates back at least to Chaum), inconsistently applied

Decoupling easiest when splitting by **entity** and **mechanism**

➔ E.g. split **authentication** from **connectivity**

Applying decoupling is protocol / context specific

Context

Ordinary data confidentiality is nearly solved

➔ TLS is everywhere, data is encrypted at rest, etc.

What remains is a layered metadata privacy problem

➔ Many overlapping solutions needed

Privacy challenges are fundamental to the Internet

➔ We rely upon others to carry our traffic/process our requests

A Bit of Terminology

Let's define sensitive / non-sensitive information:

- ➔ ▲ = sensitive user identity
- ➔ △ = non-sensitive user identity
- ➔ ● = sensitive user data
- ➔ ⊙ = non-sensitive user data

Tuples describe knowledge of some party in some context:

- ➔ E.g. (▲,⊙) = sensitive user identity + non-sensitive user data

Caveats

Identity and data are always shades of gray

- ➔ Difficult to cleanly categorize as sensitive/non-sensitive
- ➔ Identity and data are sometimes mixed/conflated
- ➔ But: still useful to analyze with generally-understood categories

Example: Mix-nets / Tor

Sender

➔ Sending a message (request/data) to some receiver

➔ Trying to achieve data/metadata privacy for ID and message

Mixes

➔ Third parties relaying the data

Receiver

➔ Partially trusted party who will receive/respond to the message

Example: Mix-nets / Tor

Sender: (▲,●)

➔ All sensitive info (of course)

Mix 1: (▲,⊙), ... Mix N-1:(△,⊙)

➔ Sensitive/non-sensitive user identity + non-sensitive user data

Receiver/Mix N: (△,●)

➔ Non-sensitive user identity + sensitive user data

The

DECOUPLING

Principle

Third-parties should know at most one of: {▲,●}

Many Examples

Chaum's designs (blinded payments, mix-nets, etc.) / Tor

Privacy Pass / Private Access Tokens

Oblivious DNS

PGPP

Private Relay

Private Aggregate Statistics

Why Does This Work?

Users often care about:

➔ Hiding their (true) identity from semi-trusted services

➔ Hiding the data/metadata of their requests from untrusted parties

Users often don't care about:

➔ Whether they reveal they are a user of some public/popular service

➔ Whether they can hide a request from the service that responds to it

Cautionary Tale: Security Gateways/VPNs

Sender: (▲,●)

➔ All sensitive info

Gateway: (▲,●)

➔ All sensitive info (**problematic**)

Receiver: (△,●)

➔ Non-sensitive user identity + sensitive user data

Other Considerations

Non-collusion:

➔ Dividing knowledge between parties requires it

Hardware enclaves / TEEs:

➔ Can shift trust and thus who knows what

Side-channels:

➔ Still a problem, can change the nature of the analysis