Metadata in the pervasive context

STRINT Workshop
28 Feb, 1 Mar 2014 – London

Alfredo Pironti
alfredo.pironti@inria.fr

What is metadata?

- Data that describes other data (Wiktionary)
  - Scary! But not exactly what we mean here
- Everything but the payload of an exchanged message
  - Very broad
  - And arguably imprecise

"See you tomorrow"

- Let's be conservative

Why metadata?

- By necessity
  - Additional data with respect to the payload
  - E.g., deliver data across the network (efficiently)
    - But an address need not to *identify* the recipient!
- By nature of communication (*side channels*)
  - Function of the payload as a measurable effect
  - Time, size, pattern, power consumption, noise…
  - … put yours here! (Do we agree on a set?)
  - Interconnected, and tricky to smooth out (e.g. hardware)

Metadata is widely available

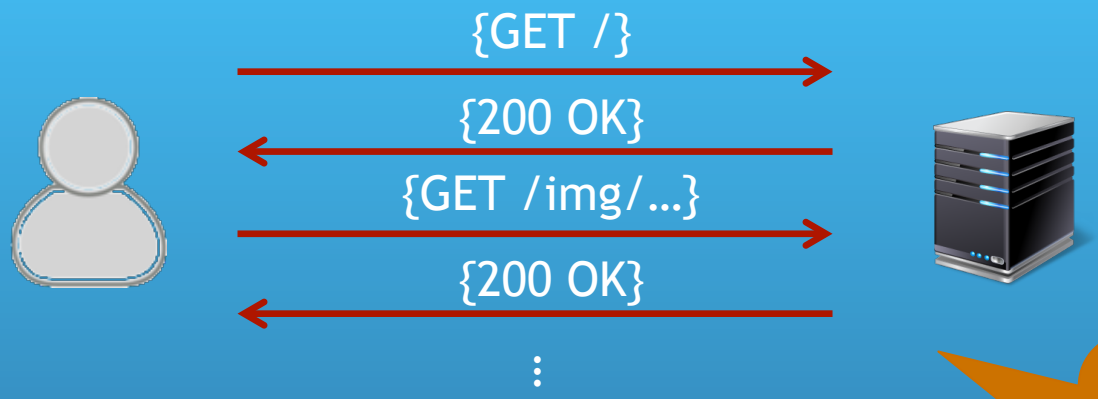
- Low-level network stack is plaintext (IP, TCP, DNS…)
  - Encrypted variants not really deployed
  - … do they properly address metadata protection anyway?
  - New standards doing worse!
    - E.g., IPv6 MAC-based addresses (50. Mike O'Neill)
- Transparent-layer crypto protocols (TLS, SSH,..) do not protect metadata by default

Transparent metadata protection?

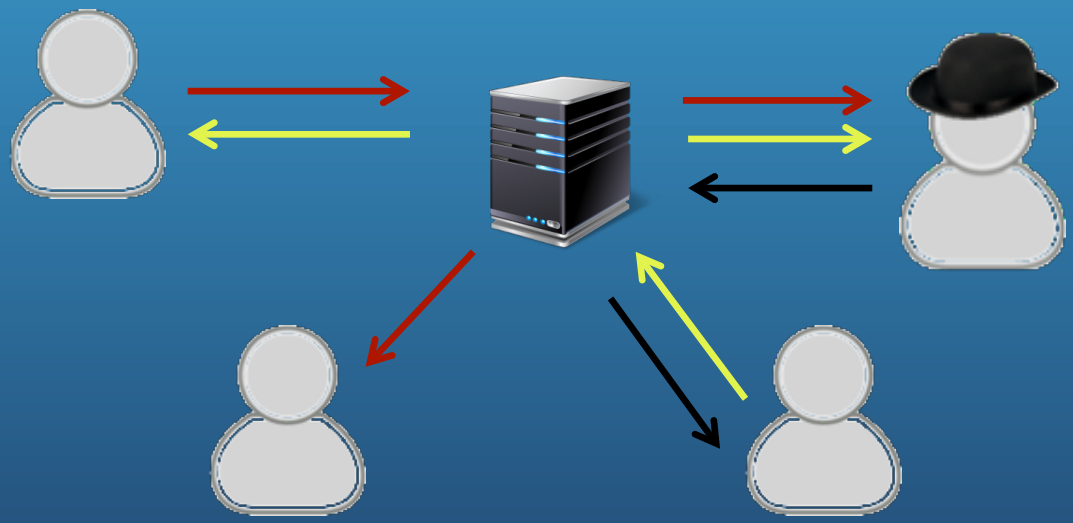
- Not really. At least in a generic and efficient way
- More optimistic with application cooperation
  - Anonymity sets, time boundaries, metadata labeling, …

Exploiting metadata

- Browsed content from browsing pattern [Chen et al., Cheng et al., Sun et al.]



- Flow of documents [Danezis, Pironti et al.]



Tor is not the definitive solution (unfortunately)

- Scales to identity fingerprint, mailing list subscription, chat …

End-to-end metadata protection

- Requires trust in the peer (54. Andreas Kuckartz)

- Users connect to many untrusted websites with valid certificates
  - CAs are not police; websites get compromised
  - Ads
  - Typically no disclosure of sensitive data; but of many metadata!

The scale of the pervasive model

- Reconstruction of social interaction
- Prediction of trends
  - Political, social, economical

(Meta)data at rest

- SaaS: data and metadata at rest on untrusted server
- Typically in the clear
  - … or encrypted under a server-controlled master key
- Even with encryption, metadata can be exploited
  - File name, size, modification time, …