CyBOK: Privacy and Online Rights Knowledge Area

Tariq Elahi
Cyber Security, Privacy, and Trust Group
School of Informatics – University of Edinburgh

Overview

- Motivate Online Privacy
- Lenses on Privacy: Considering the dimensions
- Data Privacy
- Meta-data Privacy

What is the problem?

What is the problem?

CyBOK

- Data Sharing
  - Filling out a Health questionnaire at the doctor's office
  - Revealing the location and time of your awesome house party to your close friends
  - Contextual: the same information in another setting may be privacy leak
    - Likely **aware** of the extent and parties involved

What is the problem?

CyBOK

- Data Collection
  - As a consequence of activity or interaction
  - Browsing at a coffee shop
    - MAC address recorded for profile building on store visits
    - DNS information about website visits
    - Paid for coffee or snacks (name and financial information)

- Likely **unaware** or uninformed about extent or parties involved

What is Privacy?

- About People
  - Identifying information leading to a natural person
    - Linking people to their
      - beliefs,
      - circumstances,
      - associations,
      - behaviours,
      - and more

Defining Privacy

CyBOK

- Privacy as a human right

  "No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation"

  -Article 12, UN Declaration of Universal Human Rights.

- More articulations:
  - "Information self-determination"
  - "the right to be let alone"
  - "the freedom from unreasonable constraints on the construction of one's own identity"

Privacy as…

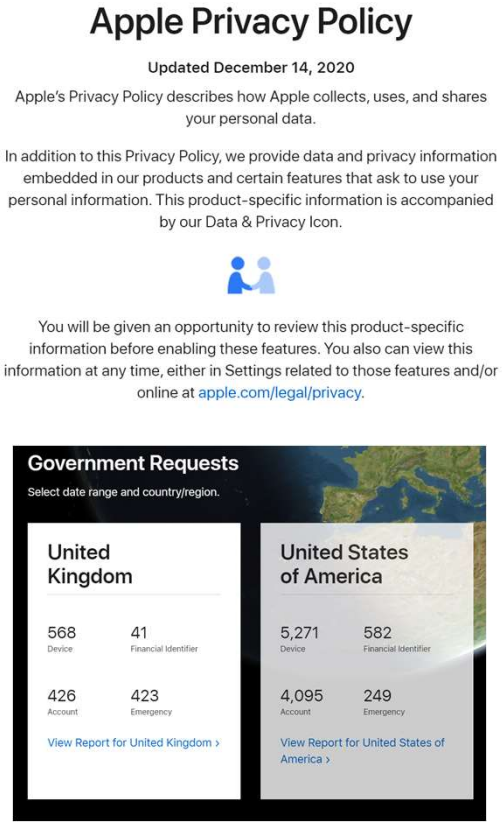
- Transparency

- Control

- Confidentiality

Privacy as Transparency

- Inform
  - what is collected and
  - For what purposes
  - By whom
  - For how long
  - and so on
- Privacy policies
- Transparency reports



Apple Privacy Policy

Updated December 14, 2020

Apple's Privacy Policy describes how Apple collects, uses, and shares your personal data.

In addition to this Privacy Policy, we provide data and privacy information embedded in our products and certain features that ask to use your personal information. This product-specific information is accompanied by our Data & Privacy Icon.

You will be given an opportunity to review this product-specific information before enabling these features. You also can view this information at any time, either in Settings related to those features and/or online at apple.com/legal/privacy.

Government Requests

Select date range and country/region.

United Kingdom

568 Device
41 Financial Identifier
426 Account
423 Emergency

View Report for United Kingdom ›

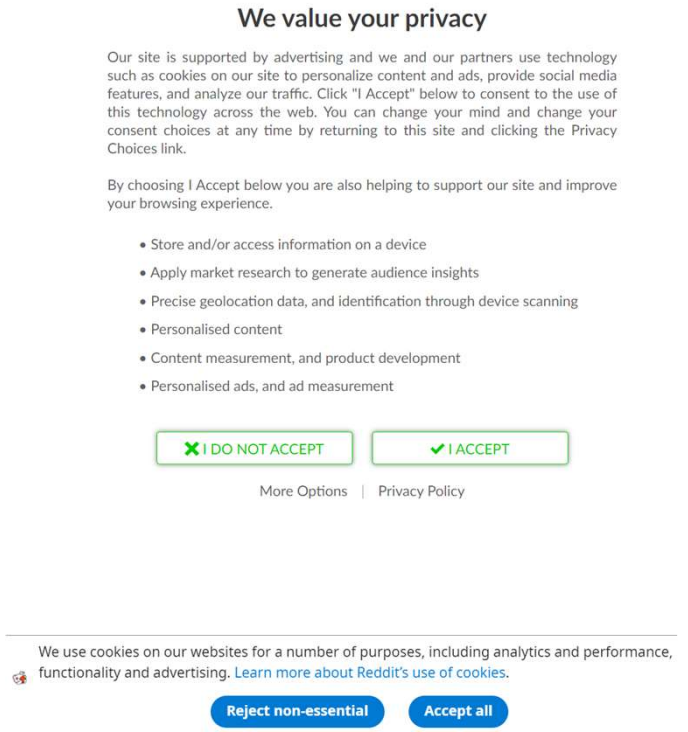
United States of America

5,271 Device
582 Financial Identifier
4,095 Account
249 Emergency

View Report for United States of America ›

Privacy as Control

- Control
  - If data can be used/shared
  - What kind of access is granted
  - What kind of processing can be done
- Consent forms



Limits of control and transparency

CyBOK

- No way to prevent (control) what is revealed
  - IP addresses on network traffic
- Explicit list of information (transparency) may not be possible to enumerate
  - Joining two or more sources of data together to form a more complete profile
    - Ad trackers
  - Unexpected correlated information in collection
    - Genomic data of family members

Privacy as confidentiality

- Hiding your information
  - Encryption
    - Secret key needed
  - Obfuscation
    - No cryptographic secrets
    - Know where to look and how to undo the obfuscation

Privacy Threat Landscape

CyBOK

- **Data** (at rest, in transit, or in processing)
  - Posts, status updates, tweets, pictures, check-ins
  - Genomic
  - Location (check-ins, maps)

- **Meta-data**
  - Timing, message lengths, IP addresses
  - Device/application fingerprints

Data Privacy (confidentiality)

Cryptography-based Access Control

- Protecting data during transit
- Protecting data during processing

Inference Control

- Anonymization
- Generalization
- Suppression
- Perturbation

Pretty Good Privacy (PGP)

CyBOK

- It primarily protects the contents of emails
- Using public-key cryptography to provide:
  - Encryption of the message (confidentiality)
  - Digital signature on the message (authenticity)
- Alice and Bob each need:
  - Private decryption key
  - Public encryption key
  - Private signing key
  - Public verification key

CyBOK

- To have a private conversation on the Internet, Alice and Bob just need to:
  - Swap their public key material (as securely as possible)
  - Then simply send encrypted messages with signatures back and forth
- They lived happily ever after (?)

What if…

CyBOK

- The adversary records all of the encrypted communication between Alice and Bob (but can not yet read anything)
- Some time later, Bob's laptop is stolen or compromised
- Bob's private keys are now exposed!
  - decryption key ← ***especially this one***
  - signing key

Key exposure

CyBOK

- Adversary can now
  - Decrypt all past messages (that were recorded between Alice to Bob)
  - Learn what *was* said by Alice
  - Alice is now identified as a sender of particular messages
  - Cryptographic evidence
    - mathematical proof since Alice's verification key verifies the signature on the messages she sent to Bob
  - Private conversations exposed after the fact

Where is the source of the problem?

CyBOK

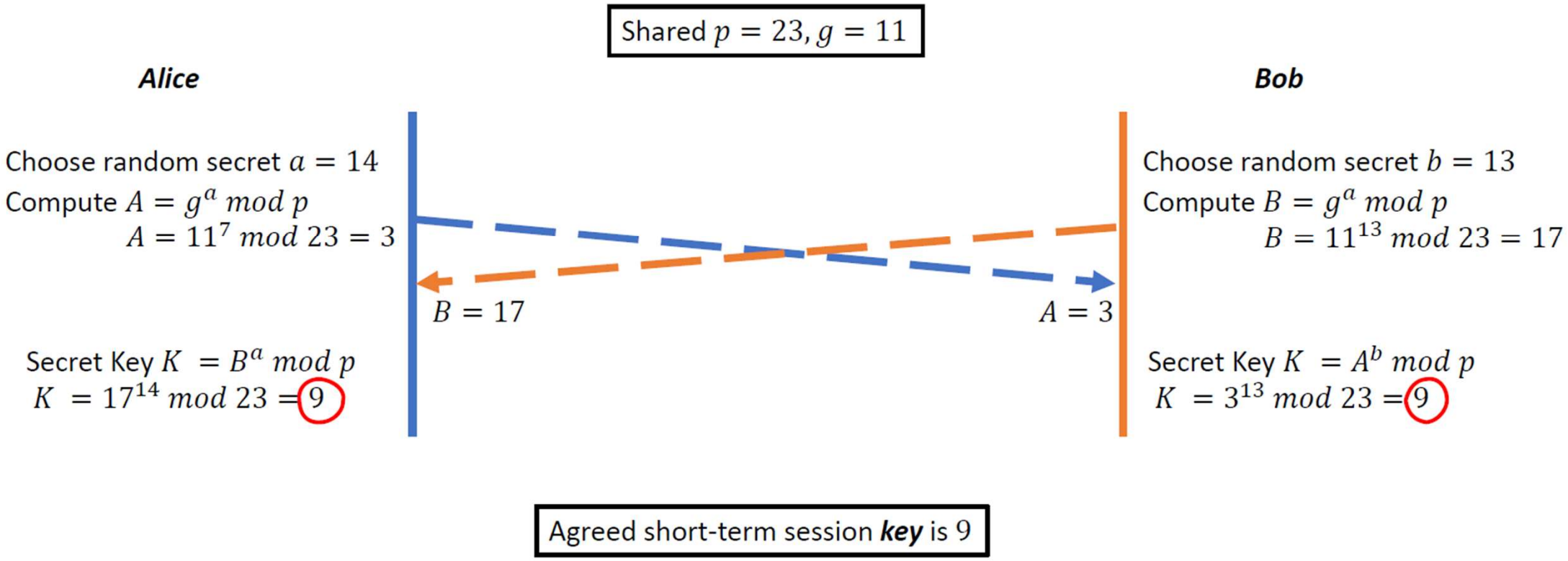
- Stolen or infected computers?
  - This problem seems hard to solve
    - You and everyone you talk to needs to never lose their laptop, click on a dubious link, install a trojan…
    - Not a good basis for trust
- PGP creates an evidence trail
  - The keys that can decrypt messages and verify signatures and provide non-repudiation
  - In an environment where keys can be exposed, Alice has to be careful about what she says
  - We'll try to solve it by fixing this part

Perfect forward secrecy

CyBOK

- Key compromised in the future should not expose past messages
- Use short-lived session (or ephemeral) keys computed from long-term keys
  - Using Diffie-Hellman key exchange
  - Session key discarded (deleted) after use
- Long-term keys used only to authenticate the DHKE messages

Diffie-Hellman key exchange



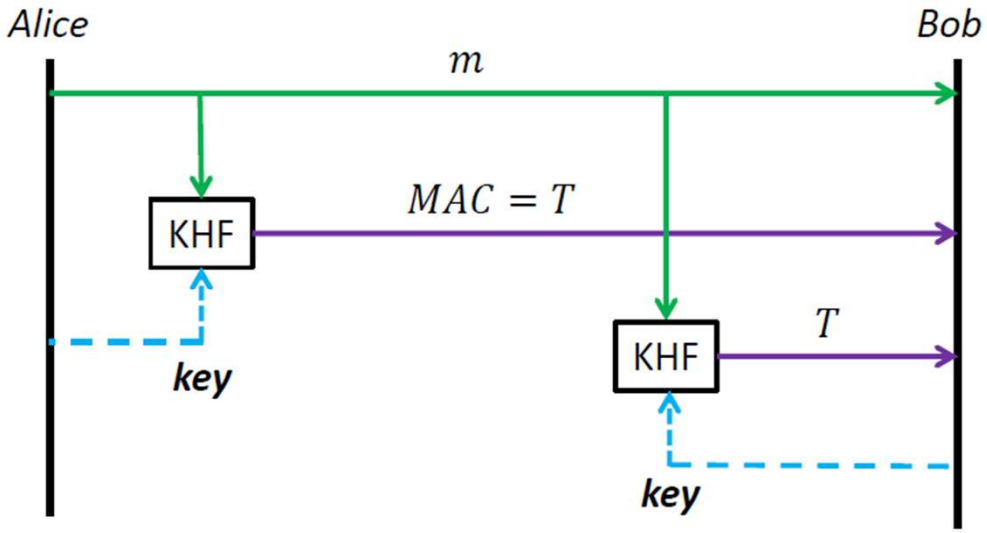
Deniable authentication

CyBOK

- Digital signatures provide *non-repudiation*, which is exactly what we do not want in this case
- Authentication is still necessary
  - Or the adversary could impersonate our friends
- We can use Message Authentication Codes (MACs) here

MACs and deniable authentication protocol

- We can create a MAC with a *keyed* hash function (KHF)
  - H(***key***, message) = MAC
  - Without the key the correct MAC can not be computed
  - *Anyone* with the key can compute the correct MAC



If $KHF(key, m) = T$ then *Alice* is the sender since only someone with the same ***key*** could produce the same T for this m

No 3rd party proof for off-the-record communication

CyBOK

- The same ***key*** was agreed between Alice and Bob
- Only that ***key*** can produce that same MAC on that message
- Bob can not prove Alice sent the message
  - Anyone with the ***key*** can also forge a new message and correct MAC
- Alice has plausible deniability

Privacy during data processing

CyBOK

- We would like to keep our data private (not share) but also use it to do useful computations in collaboration with others

- Is that even *possible*?

  - Secure multi-party computation (aka MPC or SMC)

  - Homomorphic Encryption

Desirable properties

CyBOK

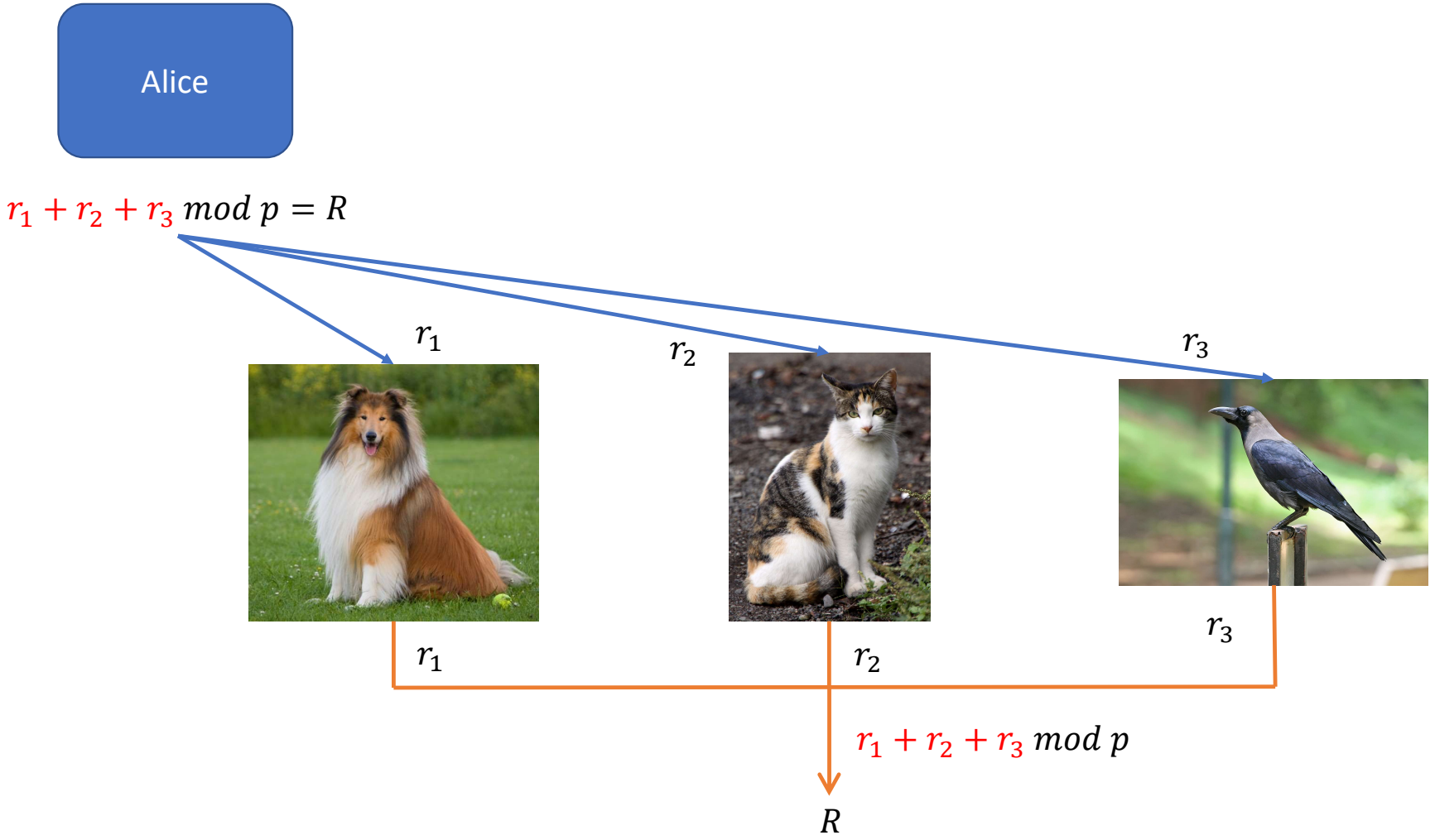
- No information about the *private inputs* is leaked
- Only the result of the computation is revealed
- What this does not mean:
  - If the function is *invertible* it can not prevent a party from learning something about the private input of other parties
  - That is not a break, it is up to the participants which functions they participate in executing
    - SMC(sum, {1,2,})=3, then any party can learn the sum without their own input

Sharing Secrets

CyBOK

- We can split a secret up in such a way that
  - An individual piece does not reveal anything about the secret
  - All pieces are required to extract the secret, otherwise nothing is revealed
    - There are ***threshold*** variants where only ***k-out-of-n*** pieces are required
  - Using simple *modular arithmetic*

CyBOK



CyBOK

Alice

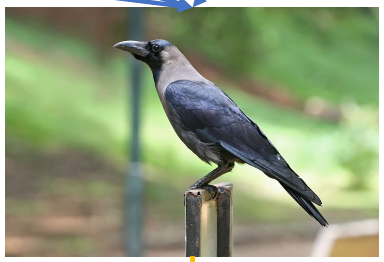
Mad Hatter

Cheshire Cat

$r_1 + r_2 + r_3 = R$

$u_1 + u_2 + u_3 = U$

$v_1 + v_2 + v_3 = V$



$S_1 = r_1 + u_1 + v_1$

$S_2 = r_2 + u_2 + v_2$

$S_3 = r_3 + u_3 + v_3$

$S_1 + S_2 + S_3$

$Sum(R + U + V)$

All additions $mod\ p$

Inference control

CyBOK

- Number-theoretic controls are restrictive on what can be done and learned
- They are inefficient for large number of participants and/or data
- More light-weight ***obfuscation-based*** inference control schemes can provide solutions where other schemes are impractical
  - Obfuscation is necessarily less secure, since it only *limits* the information leakage
  - A more *relaxed definition* of confidentiality

Obfuscation-based inference control

CyBOK

- Data Anonymization
  - k-anonymity, l-diversity, and t-closeness among others
- Data Generalization
  - Reduce the precision or bucketize the values of attributes (columns in table)
- Data Suppression
  - Do not reveal part of the dataset (delete fields, possibly based on their value)
- Dummy data addition
  - Add dummy rows to reduce the accuracy of the adversary's inferences
- Perturb the data
  - Add noise to the data to reduce the adversary's inferences

Formal approach to inference control

CyBOK

- The inference control techniques so far were ad-hoc and it is difficult to evaluate the level of privacy and utility that is achieved

- *Differential privacy* is a formal noise addition mechanism that provides *worst-case* privacy guarantees and utility trade-off

- Two variants
  - Local: noise is added before it is added to the database
  - Global: noise is added to the answer of the query and there is a DB request handler that mediates between the query requestor and the database
    - Contrast the case where the database is sanitized and then published publicly, as before

ε-differential privacy

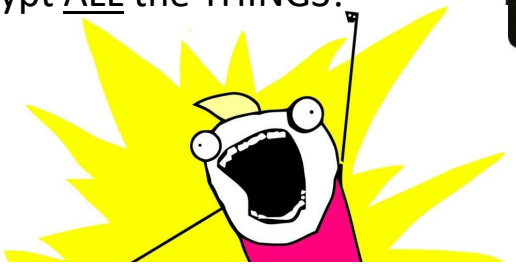
CyBOK

- Depends on the sensitivity s (the maximum difference one record makes to the query results)

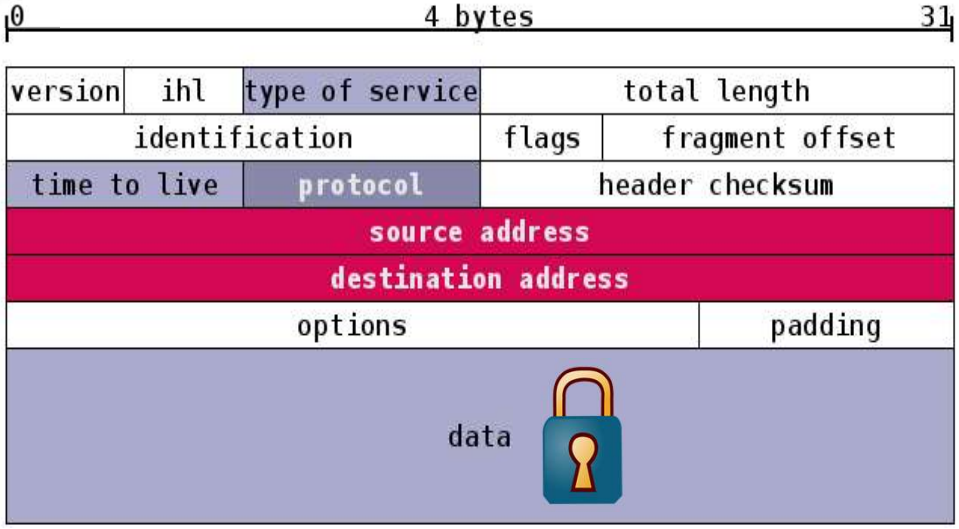
- We want to add enough noise such that it is difficult* to tell whether that record was present in the calculation of the output or not

- The noise addition is (ideally) realized with *Laplace distribution*
  - More practically, other distributions are the *Normal distribution* (however then we get (ε, δ)–differential privacy, a looser privacy bound)

Encrypt ALL the THINGS!



CyBOK



Meta-data

- Context
  - Time
  - Duration
  - Length
  - Who

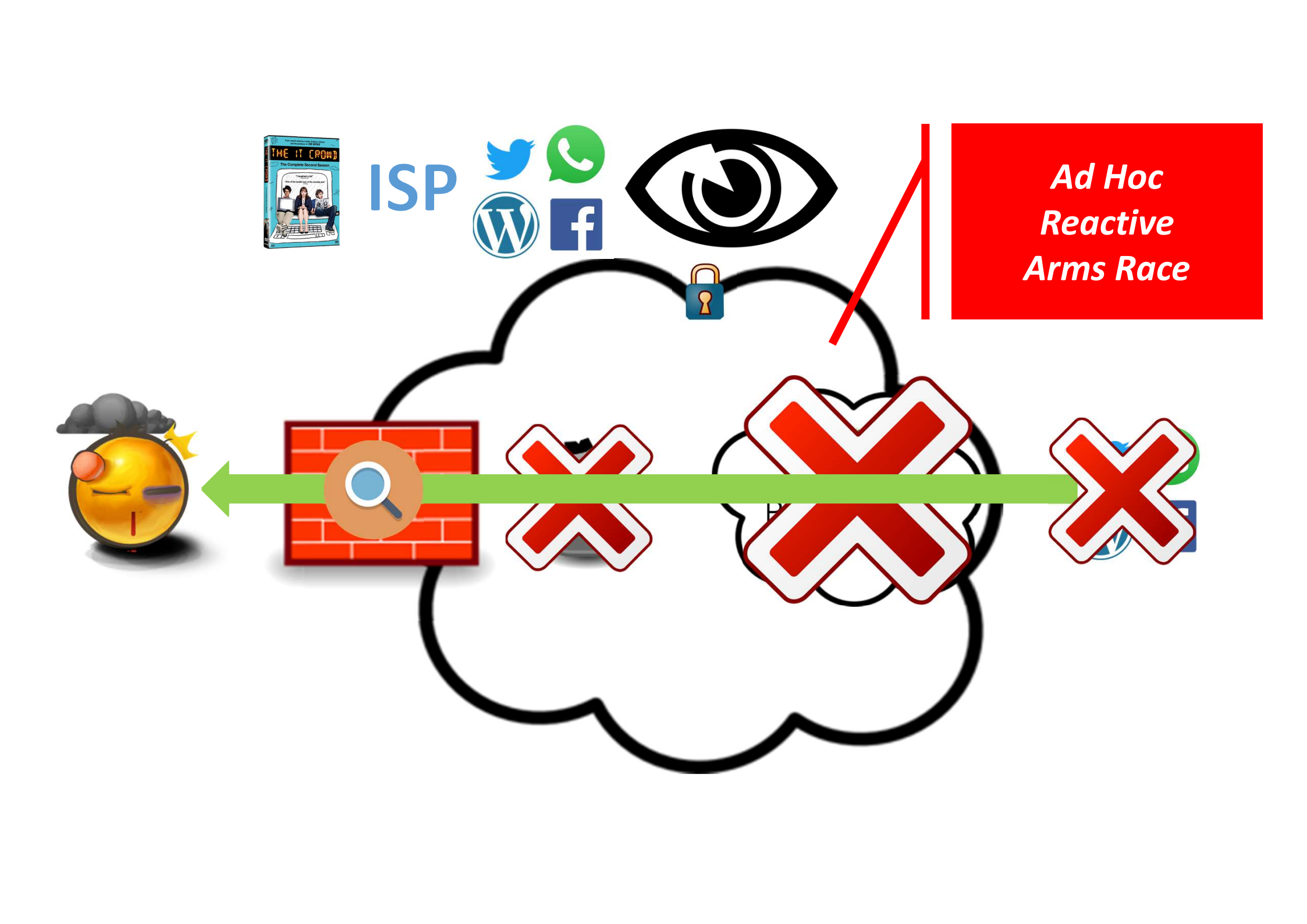
Meta-data Privacy

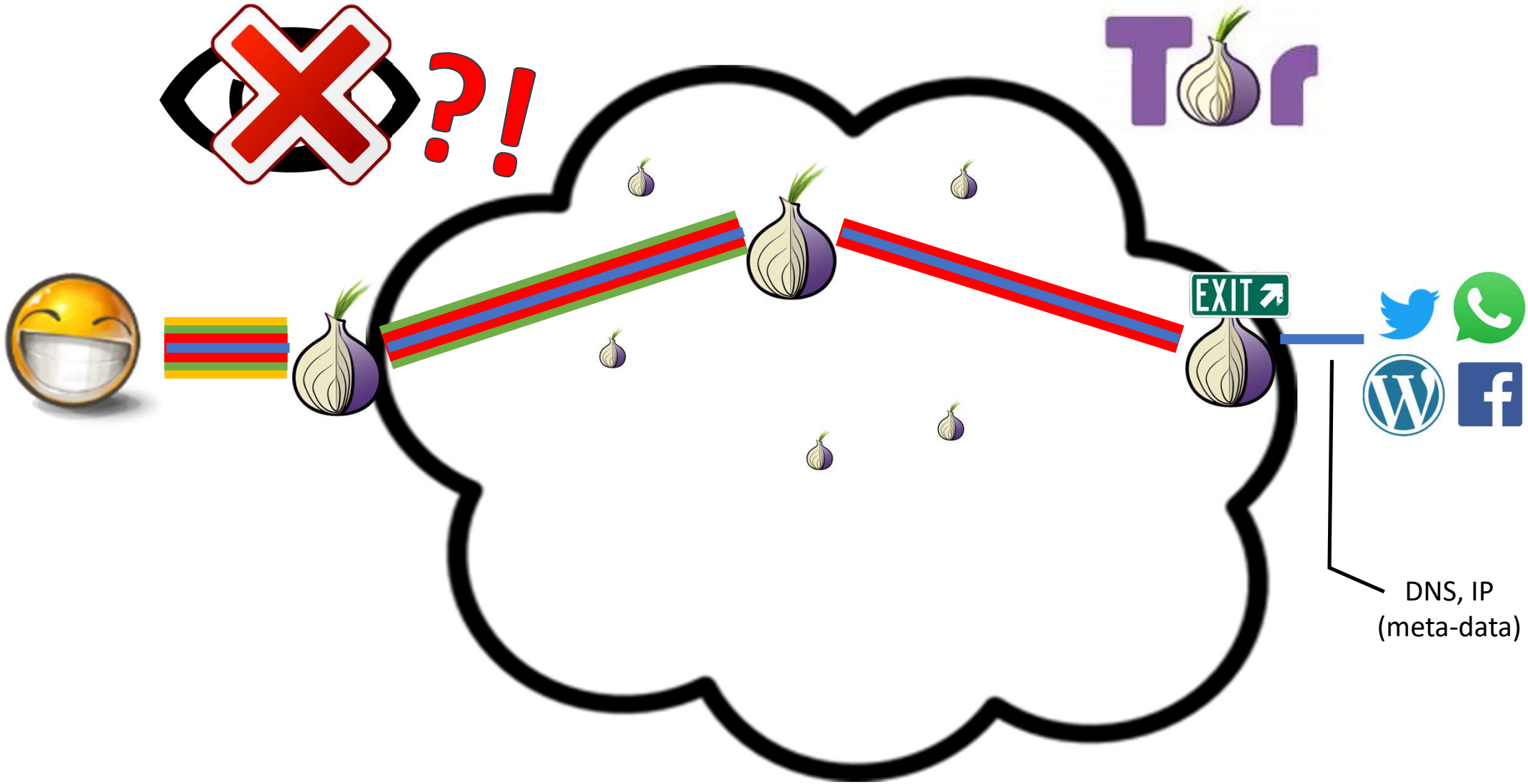
CyBOK

- Information needed to perform critical functions
  - Routing a message
  - Fastest way from the pizza place to your house
- Often the consequence of insecure design
  - Security and Privacy not part of the initial design of the system
- Provides an avenue for unwanted information gathering

ISP

CyBOK


THE IT CROWD
The Complete Second Season
ISP
Ad Hoc
Reactive
Arms Race


?!
Tor
EXIT
DNS, IP
(meta-data)

Some attacks on Tor

CyBOK

- End-to-end correlation attack
  - Timing attacks
- Selective denial of service
  - Path bias
- Website fingerprinting

Mix networks

CyBOK

- Reordering messages removes the timing information that an adversary can use to track a message
- Low-latency systems (like Tor) are susceptible to timing attacks
  - Especially from a global adversary
- Mixing well can remove this vector of attack
  - Cost of delay
  - The more messages to mix with the higher the anonymity set

Mixing types

- Timed – Buffer messages for a set time interval and then send them all out.
- Threshold – Buffer messages until a threshold number of messages is reached and then send them all out.
- Pool – Buffer messages until there are at least a threshold number and then send out only a fraction of them every set time interval.
- Continuous - each message is delay independently according to the delay chosen by the sender. Delays sampled from exponential distribution.

Expectation on delays

CyBOK

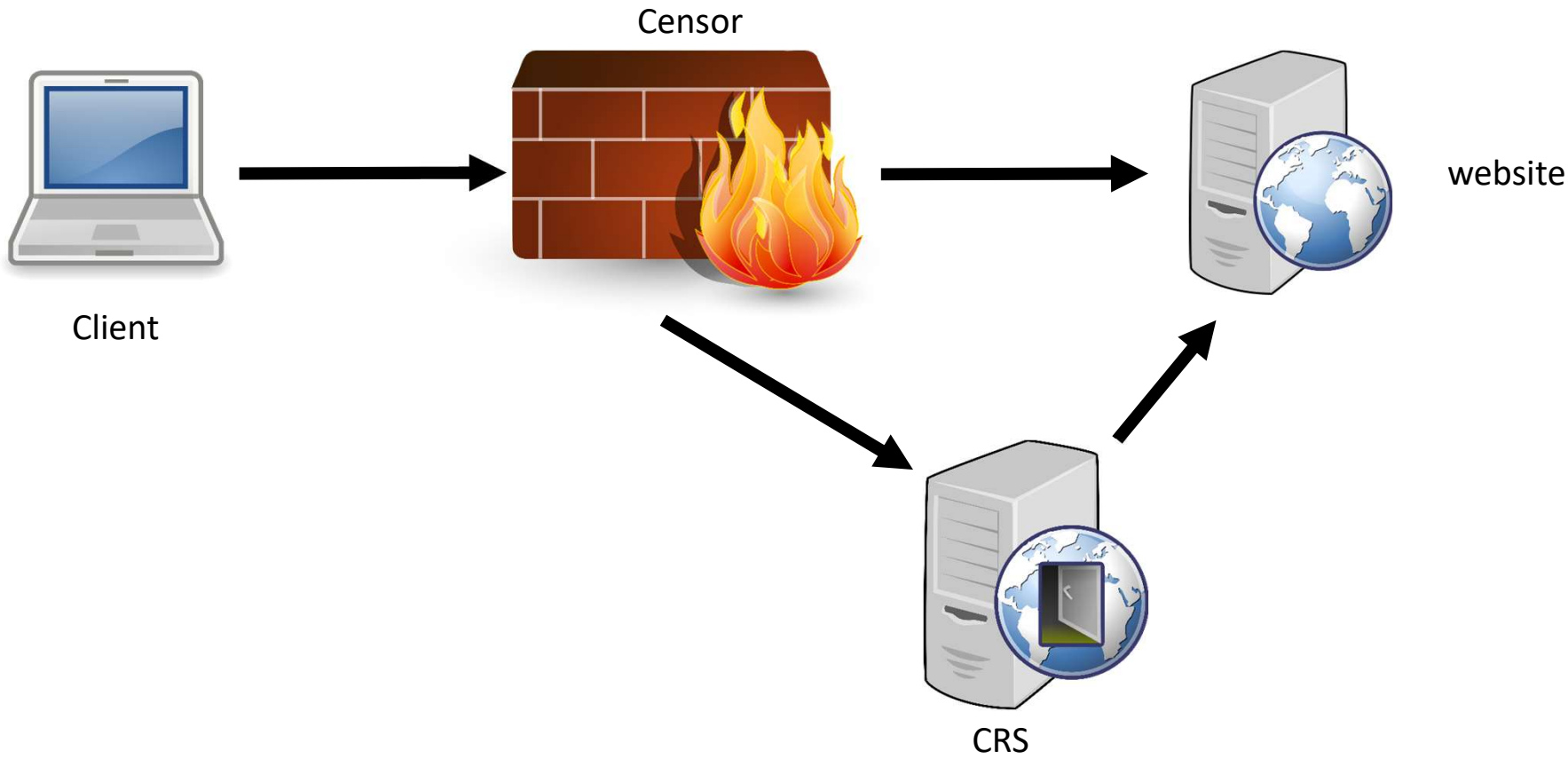
- Timed mixes provide a known message delay
  - But variable level of mixing
- Threshold mixes provide a known amount of mixing
  - But variable expected delay
- Pool mixes provide a lower bound on the amount of mixing
  - Delay is now an expectation based on the emission fraction
- Continuous mixes provide an expectation on both the mixing and delay
  - These depend on the sending rate of the client population and the delay parameter

Internet Censorship

CyBOK

- Governments, Corporations, Service Providers
- Prevent information flow
  - Prevent the **publication** of content
  - Prevent the **access** to content
  - Assumption: rational actor desiring net positive utility
    - Do not want to block the Internet as a matter of course unless it is more cost effective

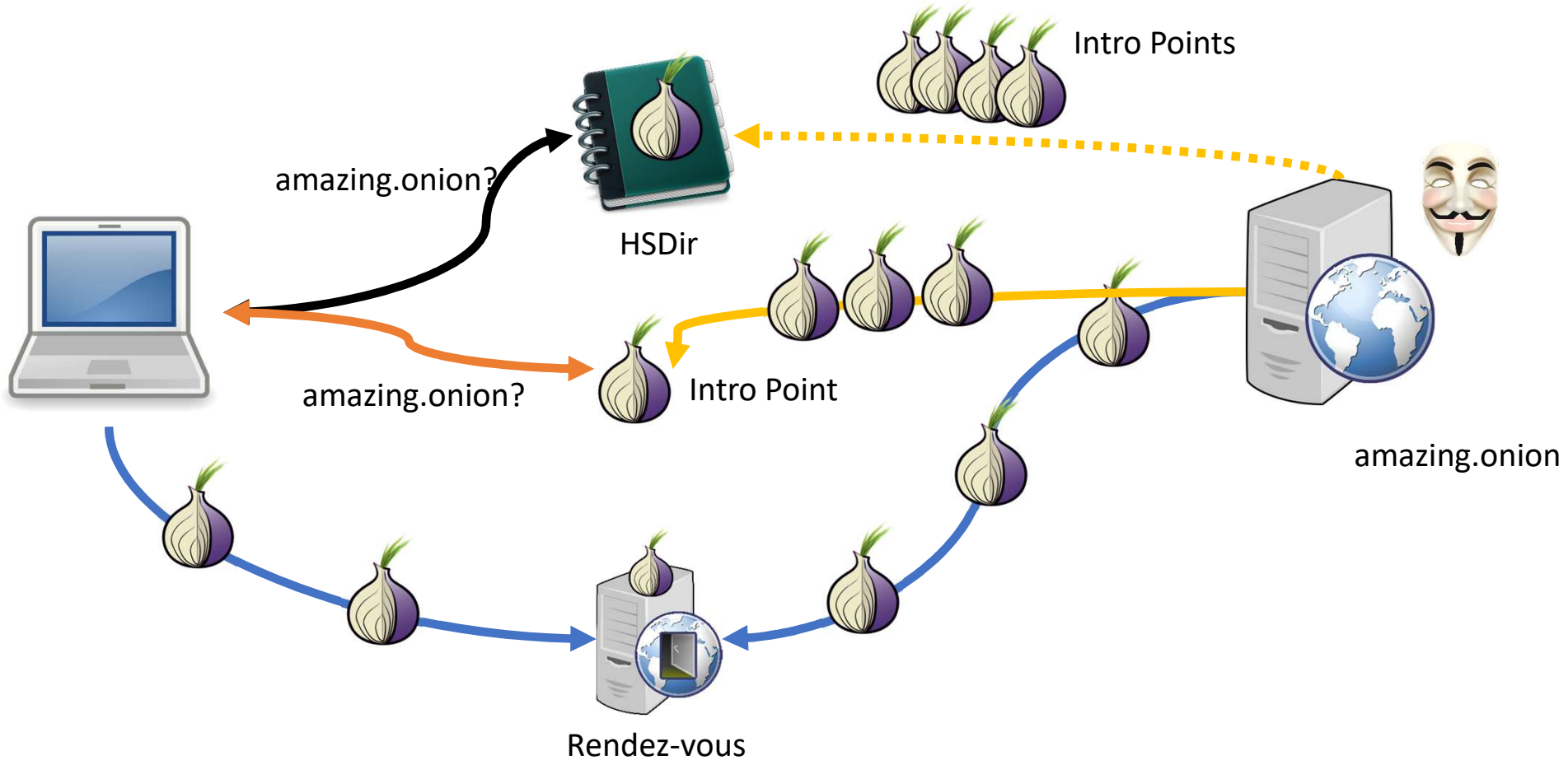
Censorship Resistance Systems in a Nutshell



Common Censorship Resistance Strategies

- No Single Point of Failure
  - Store multiple copies of content, add multiple access paths/points, spread over different jurisdictions, regions, and operators
- Collateral Damage
  - Hide censorship traffic amongst allowed traffic. Any interference with the censored traffic will also impact unrelated traffic. Used as a deterrent.
- Do Not Look Suspicious
  - Look different **enough** from not allowed usage or look **exactly** like allowed usage
- Be Untraceable
  - Do not present a target (e.g. hidden IP address)

Example – Tor Onion Services



Key take-aways

- Privacy can be seen through the lenses of Transparency, Control, and Confidentiality
- A layered approach is advisable using both societal and technological approaches
- Data and Meta-data privacy technologies are available to enable a large number of use cases, and are practical for use
  - More advanced in certain areas, like anonymous communications
- An active arms race with the adversary and defenders in constant struggle for supremacy