EPFL

Computer Security (COM-301)

Privacy

Slides created by Carmela Troncoso
Some slides/ideas adapted from: George Danezis, Bart Preneel, Claudia Diaz, Seda Guerses

Goal of this lecture

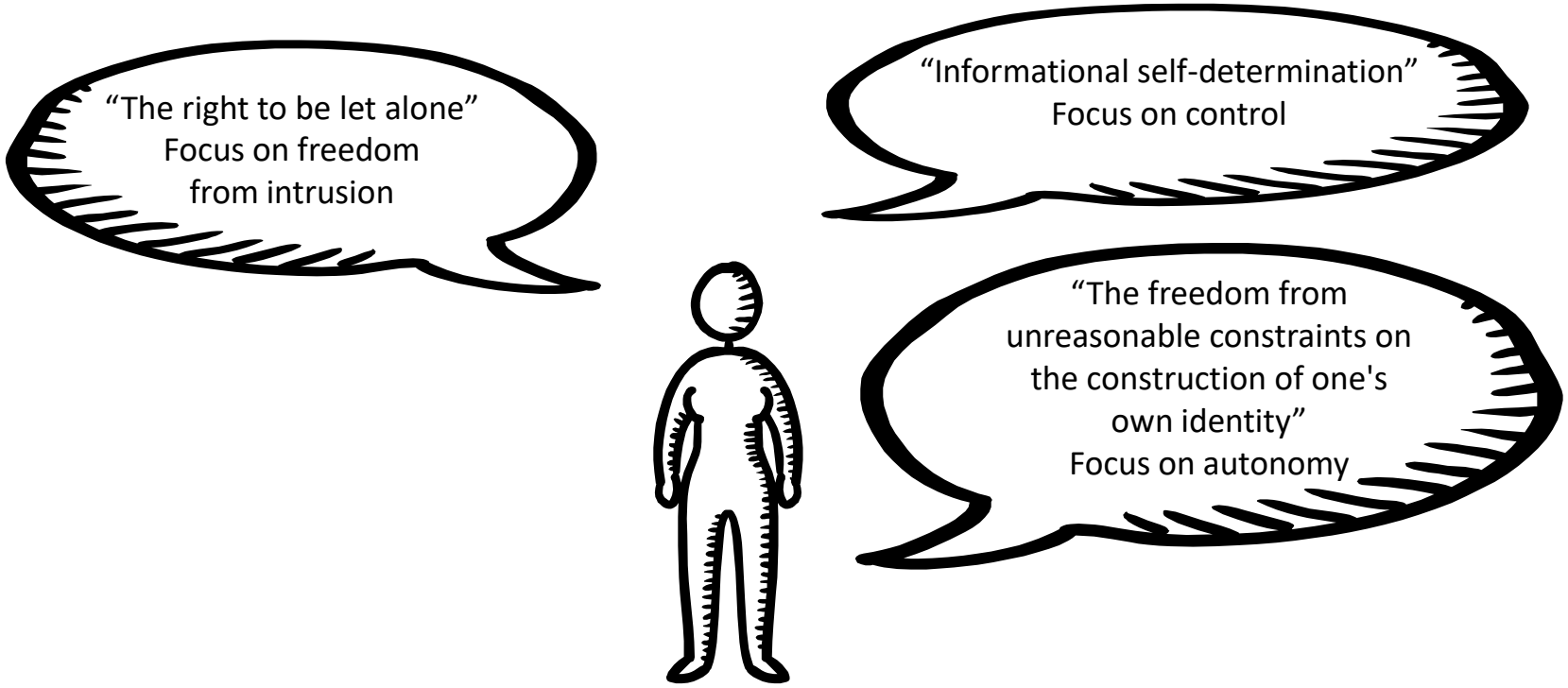
Understanding:

- Privacy is not solely a local, individual-oriented problem. It is a *global property*
- There are different conceptions of privacy depending on the adversary model
- Depending on the adversary model one relies of different Privacy Enhancing Technologies: different protection degree
- Privacy requires to protect information beyond content: The need to protect meta-data

What is privacy

Abstract and subjective concept, hard to define

Dependent on cultural issues, study discipline, stakeholder, context



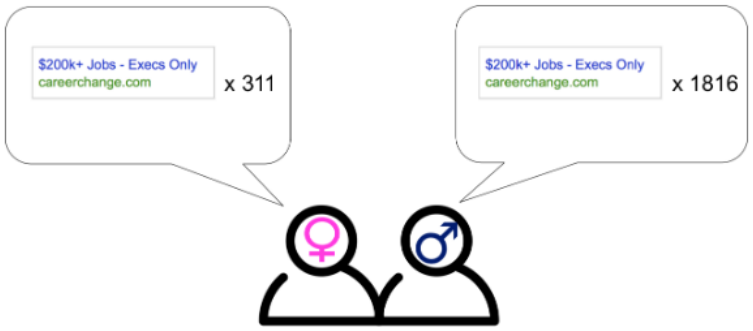
The context: Availability of data

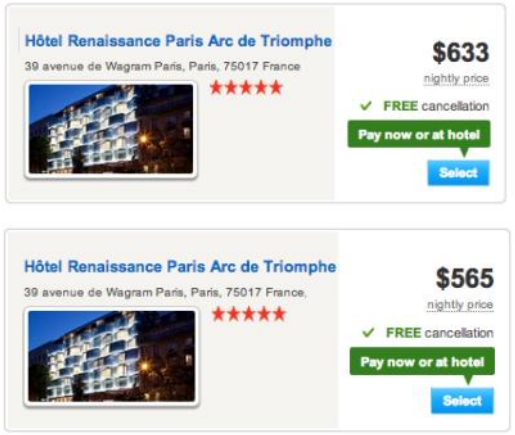
Intelligent data-based applications

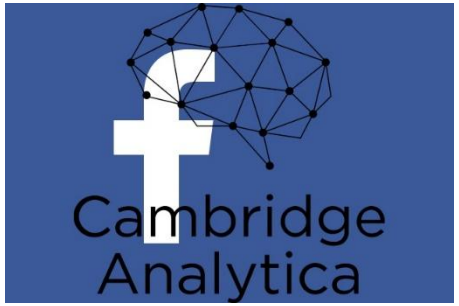
- Recommendation systems
  - Movies (Netflix)
  - Products (Amazon)
  - Friends (Social networks)
  - Music (Spotify, iTunes)
- Location based services
  - Friend finders
  - Maps
  - Points of interest
- Health monitoring
- Children/Elderly trackers
- Smart metering
- Intelligent buildings

Individual applications are legitimate

Together they become a cheap
SURVEILLANCE INFRASTRUCTURE







100K users installed CA Facebook App

enabled **COLLECTING PERSONAL DATA** of 87+ million

public profile, page likes, birthday and current city

creation of **PROFILES** of the subjects of the data

TARGETED ADVERTISEMENTS during the US elections

The context: Availability of data

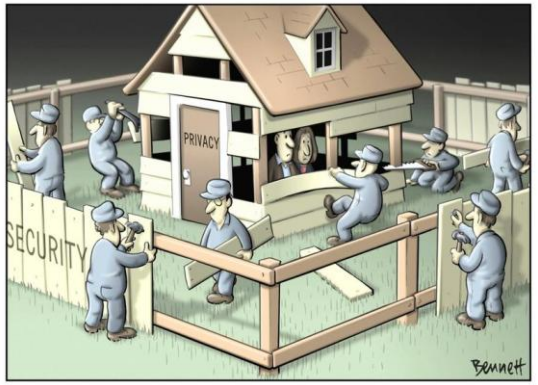
Intelligent data-based applications

- Recommendation systems
  - Movies (Netflix)
  - Products (Amazon)
  - Friends (Social networks)
  - Music (Spotify, iTunes)
- Location based services
  - Friend finders
  - Maps
  - Points of interest
- Health monitoring
- Children/Elderly trackers
- Smart metering
- Intelligent buildings

Individual applications are legitimate

Together they become a cheap
SURVEILLANCE INFRASTRUCTURE

We need privacy!



But what about security!!?!?!?!

Common belief: we need to tradeoff security for privacy!

"For National Security surveillance is good and privacy is bad"

(Surveillance == Security) == True ??

*Surveillance may be not **effective***: smart adversaries evade surveillance

criminals use Telegram, Threema, Signal,... but average users do not!!

*Surveillance tools can be **abused***: lack of transparency and safeguards

NSA spying on Americans, Spanish ministry spying independentist politicians, Companies

*Surveillance tools can be **subverted** for crime*

Greek Vodafone scandal (2004-2005): "someone" used the legal interception functionalities (backdoors) to monitor 106 key people

Privacy **IS** a security property

For individuals

protection against profiling and manipulation.
protection against crime / identity theft

For companies

protection of trade secrets, business strategy, internal operations, access to patents

For governments / military

protection of national secrets, confidentiality of law enforcement investigations, diplomatic activities, political negotiations

Privacy IS a security property

Denying privacy to some is denying privacy to all!!

INFRASTRUCTURE IS SHA

Individuals, Industry, and Governments use the same applications

Directly
(Cloud-based services, Industry 4.0, Blockchain)

Indirectly
(employers are users)

and Privacy is important for society

Daniel Solove,
Prof. of Law

"Part of what makes a society a good place in which to live is the extent to which it allows people freedom from the intrusiveness of others. **A society without privacy protection would be suffocation**"

Not so much Orwell's "Big Brother" as Kafka's "The Trial":

"...a bureaucracy with inscrutable purposes that uses people's information to make important decisions about them, yet denies the people the ability to participate in how their information is used"

"The problems captured by the Kafka metaphor are of a different sort than the problems caused by surveillance. They often do not result in inhibition or chilling. Instead, they are problems of information processing—the storage, use, or analysis of data—rather than information collection."

"...not only frustrate the individual by creating a sense of helplessness and powerlessness, but they also affect social structure by altering the kind of relationships people have with the institutions that make important decisions about their lives."

may become

ONE RING TO RULE THEM ALL

What is privacy in Privacy Enhancing Technologies

PETs

3 different types of PETs depending on ...

the concerns they address

their goals

their challenges and limitations

Gürses, Seda, and Claudia Diaz. "Two tales of privacy in online social networks." IEEE Security & Privacy 11.3 (2013): 29-37.
Diaz, Claudia, and Seda Gürses. "Understanding the landscape of privacy technologies." Information Security Summit (2012): 58-63.
Danezis, George, and Seda Gürses. "A critical review of 10 years of privacy technology." Surveillance cultures: a global surveillance society (2010): 1-16.

1 – The adversary is in your **social** circle

CONCERNS - The privacy problem is defined by **Users**

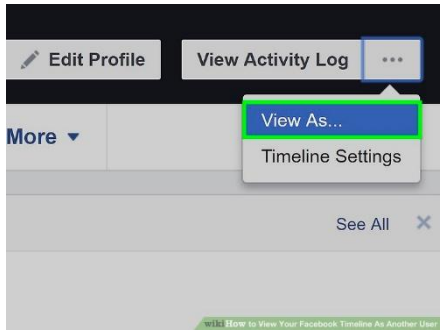
Technology brings problems

- "My parents discovered I'm gay"
- "My boss knows I am looking for other job"
- "My friends saw my naked pictures"

GOALS - Do not surprise the user

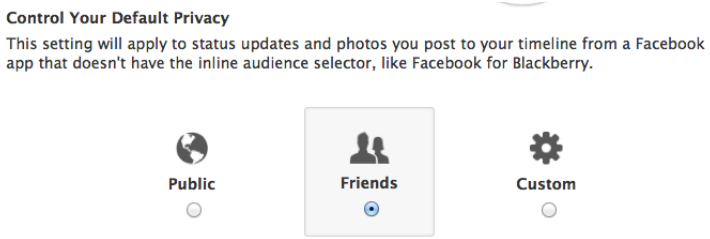
Two main approaches

- Support decision making
- Help identifying actions impact



Contextual feedback

Privacy nudges



Easy defaults

1 – The adversary is in your **social** circle

CONCERNS - The privacy problem is defined by **Users**

Technology brings problems

- “My parents discovered I'm gay”
- “My boss knows I am looking for other job”
- “My friends saw my naked pictures”

GOALS - Do not surprise the user

Two main approaches

- Support decision making
- Help identifying actions impact

LIMITATIONS

Only protects from other users: **trusted service provider**!

Limited by users’ capability to understand policies

Based on user expectations – What if the expectations are null?

1 – The adversary is in your **social** circle

CONCERNS - The privacy problem is defined by **Users**

Technology brings problems

- "My parents discovered I'm gay"
- "My boss knows I am looking for other job"
- "My friends saw my naked pictures"

GOALS - Do not surprise the user

Two main approaches

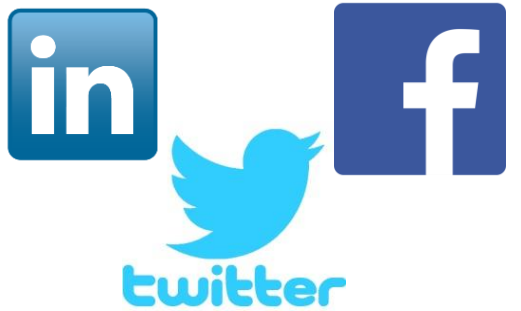
- Support decision making
- Help identifying actions impact

LIMITATIONS

Only protects from other users: **trusted service provider**!

Limited by users' capability to understand policies

Based on user expectations – What if the expectations are null?



Common Industry approach
Make users comfortable

2 – The provider may be adversarial (Institutional Privacy)

CONCERNS - The privacy problem is defined by **Legislation**

Data **should not** be collected without user consent or processed for illegitimate uses

Data **should** be secured: correct, integrity, deletion

Personal data
any information that relates to an identified or identifiable living individual.

2 – The provider may be adversarial (Institutional Privacy)

CONCERNS - The privacy problem is defined by **Legislation**

Data **should not** be collected without user consent or processed for illegitimate uses

Data **should** be secured: correct, integrity, deletion

GOALS – Compliance with data protection principles

- informed consent
- purpose limitation
- data minimization
- subject access rights

Preserving the security of data

Auditability and accountability

2 – The provider may be adversarial (Institutional Privacy)

CONCERNS - The privacy problem is defined by **Legislation**

Data **should not** be collected without user consent or processed for illegitimate uses

Data **should** be secured: correct, integrity, deletion

GOALS – Compliance with data protection principles

informed consent
purpose limitation
data minimization
subject access rights

Preserving the security of data
Auditability and accountability

Access control
Logging
Anonymization??

Wouldn't it be nice if... you could take a dataset full of personal data, and transform it into one with no personal data – while keeping all the value of the data?

Magic does not exist!

this **cannot** happen in general!

2 – The provider may be adversarial (Institutional Privacy)

CONCERNS - The privacy problem is defined by **Legislation**

Data **should not** be collected without user consent or processed for illegitimate uses
Data should be secured: correct, integrity, deletion

GOALS – Compliance with data protection principles

informed consent
purpose limitation
data minimization
subject access rights

Preserving the security of data
Auditability and accountability

LIMITATIONS

Never questions collection – assumes it is necessary
Trusted service provider! No technical measures to protect data from them
Limits misuse, but not collection (seven legal basis)
Limited scope (personal data != all data)

3 – “Everyone” is the adversary (Anti-surveillance Privacy)

CONCERNS - The privacy problem is defined by **Security Experts**

Data is disclosed **by default** through the ICT infrastructure: **the adversary is anybody**

Concerned about: censorship, surveillance, freedom of speech,…

3 – Everyone is the adversary (Anti-surveillance Privacy)

CONCERNS - The privacy problem is defined by **Security Experts**

Data is disclosed **by default** through the ICT infrastructure: **the adversary is anybody**

Concerned about: censorship, surveillance, freedom of speech,…

GOALS – Minimize

- Default disclosure of personal information to anyone - both explicit and implicit!
- Minimize the need to trust others

3 – Everyone is the adversary (Anti-surveillance Privacy)

CONCERNS - The privacy problem is defined by **Security Experts**

Data is disclosed **by default** through the ICT infrastructure: **the adversary is anybody**

Concerned about: censorship, surveillance, freedom of speech,…

GOALS – Minimize

Default disclosure of personal information to anyone - both explicit and implicit!

Minimize the need to trust others

LIMITATIONS

Privacy-preserving designs are narrow – very difficult to create “general purpose privacy”

Usability problems both for developers and users

how the @$%&#$Ŷ& do I program this?

performance hit

unintuitive technologies

Industry lacks incentives

Some Privacy Technologies

Everyone is the adversary

The adversary is almost anyone and VERY powerful

Intelligence agencies

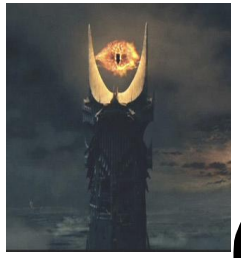
Your Parents

Your Children

Your hardware?

The Boss

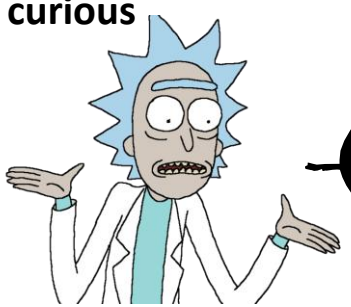
Your Roomates



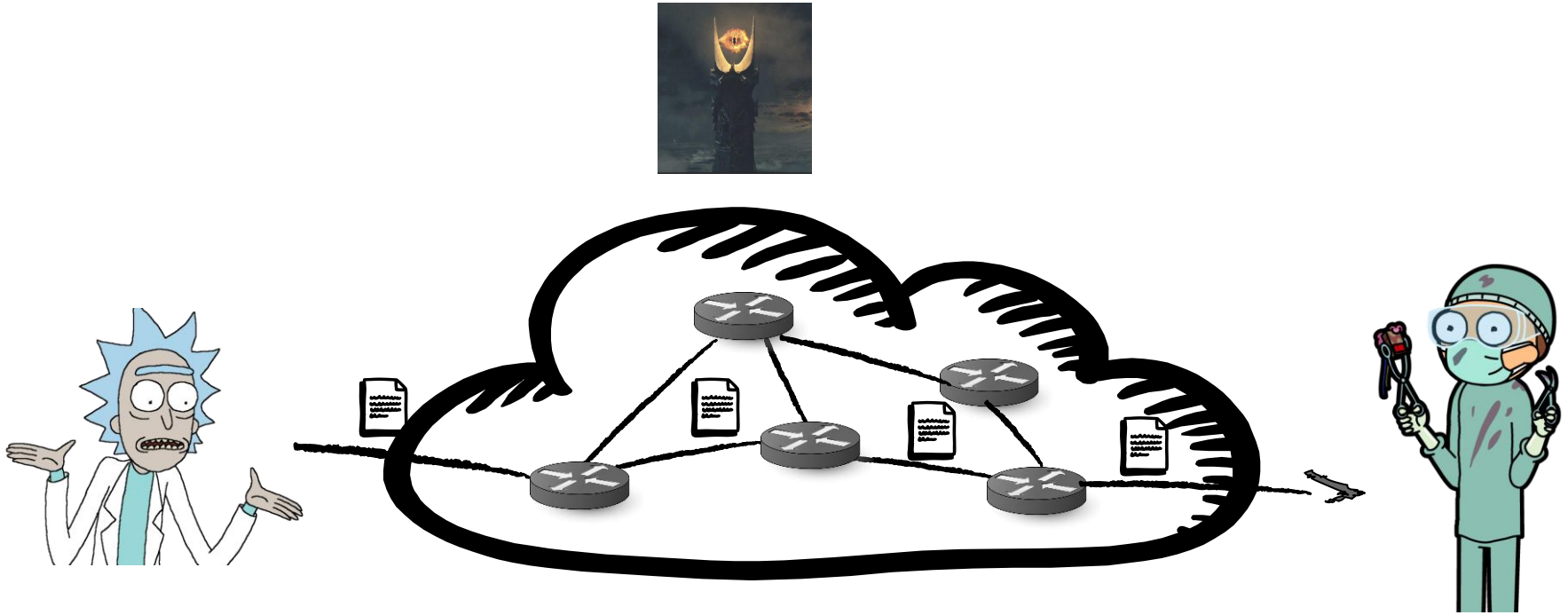
Anybody curious

ISPs

SysAdmins

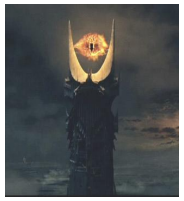


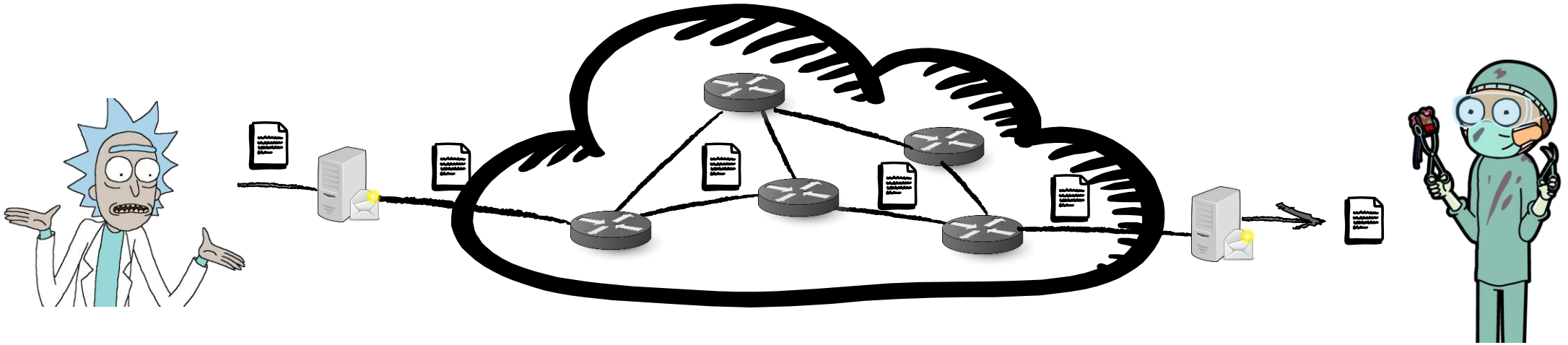
End to End Encryption



End to End Encryption

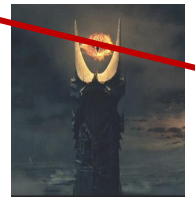
What is an End?

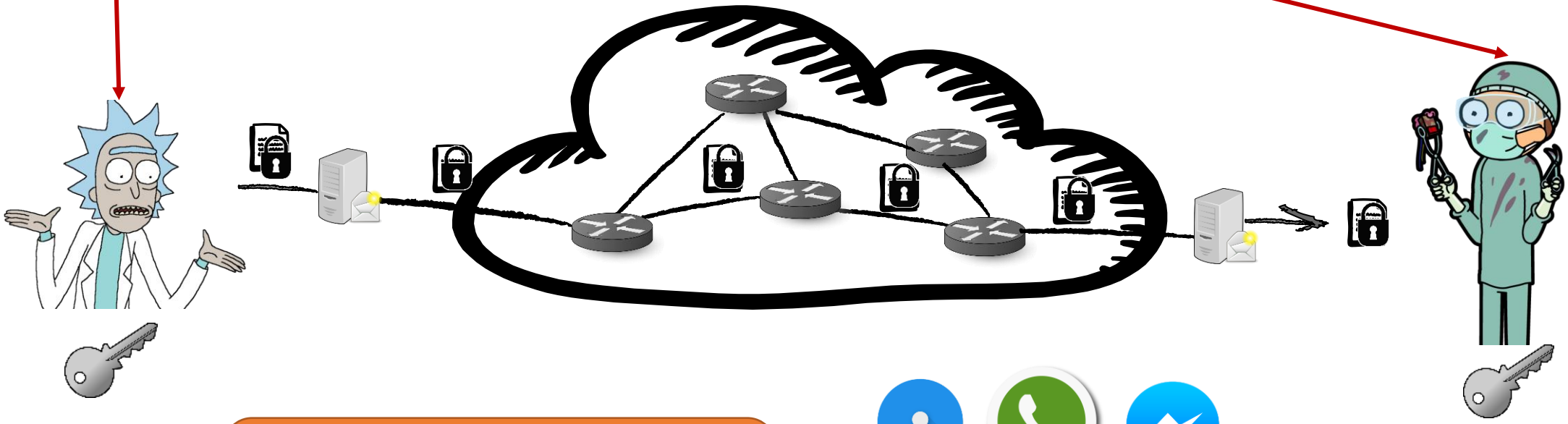




Cryptography → Confidentiality!
(and integrity and authenticity)

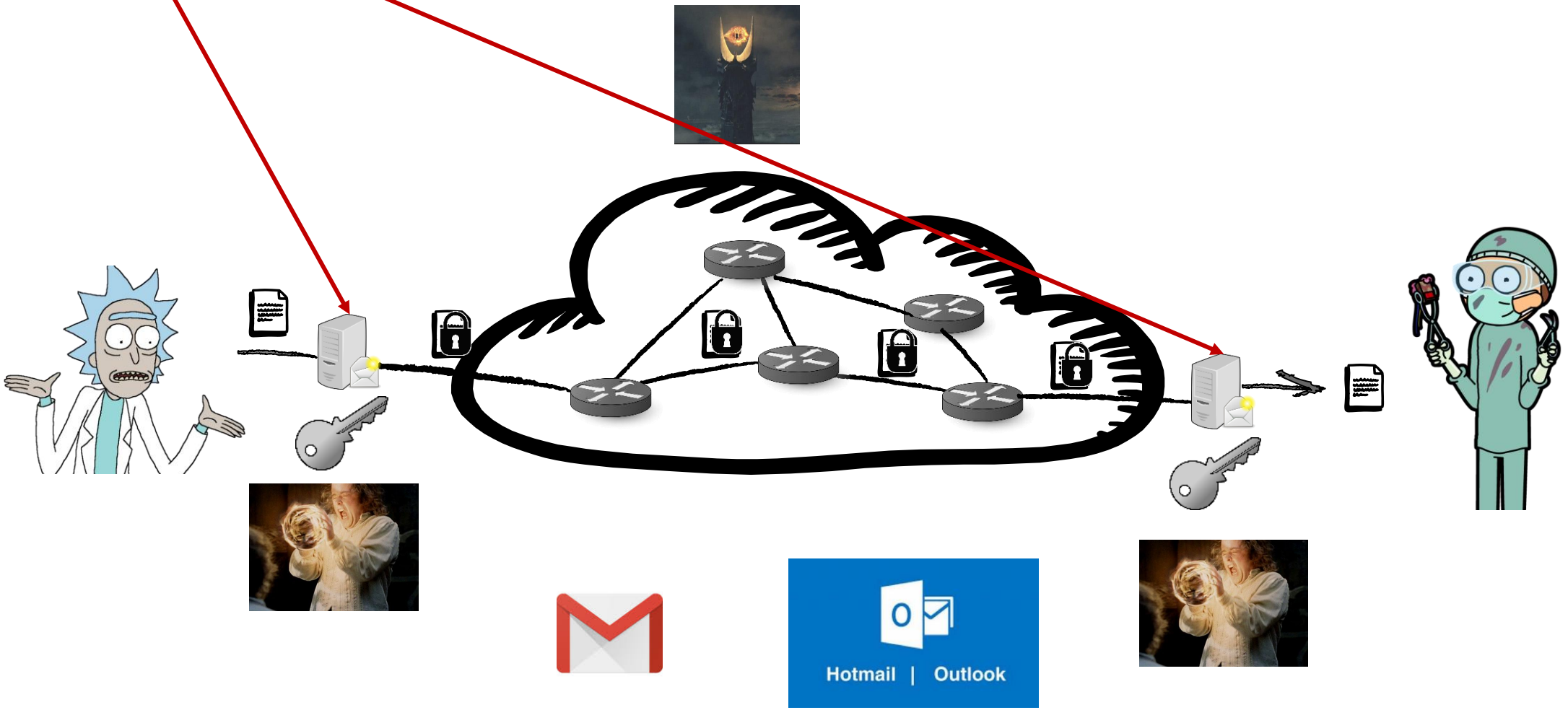
End to End Encryption



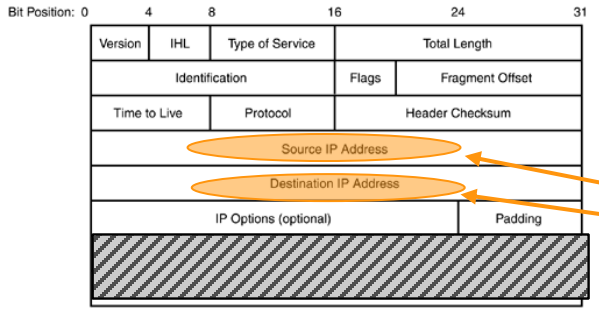


They also provide forward secrecy,
by using ephemeral keys.
How: advanced crypto

End to End Encryption



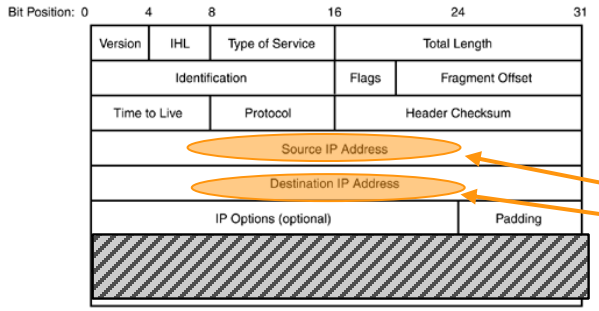
But we can encrypt! What is the problem?



IPv4 Header (RFC 791, 1981)

Same for Ethernet, TCP, SMTP, IRC, HTTP, ...

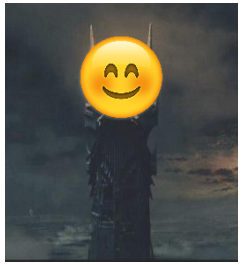
The problem is Traffic Analysis



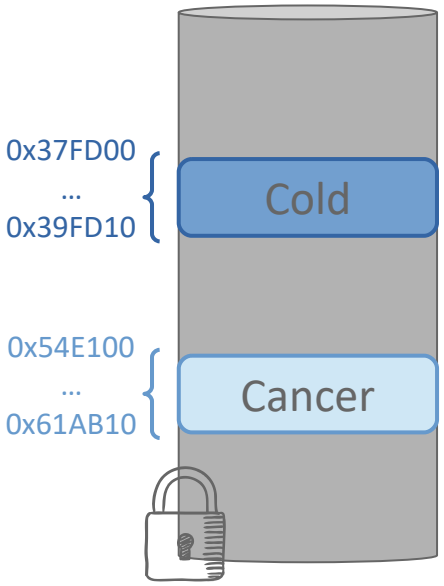
IPv4 Header (RFC 791, 1981)

Same for Ethernet, TCP, SMTP, IRC, HTTP, ...

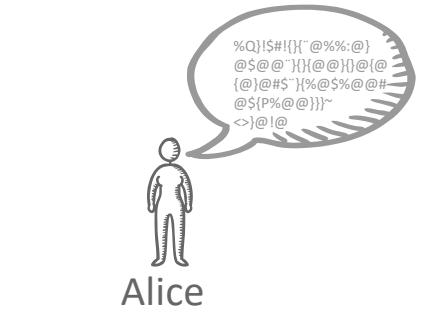
Other metadata is also sensitive!!



Implicit data is as important as explicit data!



The address where data is stored may reveal information about the content.
Example: medical database with patients with mild and severe diseases in different locations



The address where an action happens may reveal information about the action / user.
Example: sending a message from an Oncologist clinic reveals information about the sender

Traffic analysis

Wikipedia: traffic analysis is the process of intercepting and examining messages in order to deduce information from patterns in communication

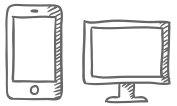
Making use of "just" traffic data of a communication (aka metadata) to extract information (as opposed to analyzing content or perform cryptanalysis)


Identities of communicating parties


Timing, frequency, duration


Location


Volume


Device

MILITARY ROOTS

M. Herman: "These non-textual techniques can establish **targets' locations**, order-of-battle and **movement**. Even when messages are not being deciphered, traffic analysis of the target's Command, Control, Communications and intelligence system and its patterns of behavior provides indications of his **intentions** and **states of mind**"

WWI: British troops finding German boats.

WWII: assessing size of German Air Force, fingerprinting of transmitters or operators (localization of troops).

NOWADAYS

Diffie&Landau: "Traffic analysis, not cryptanalysis, is the backbone of communications intelligence"

Stewart Baker (NSA): "metadata **absolutely tells you everything about somebody's life**. If you have enough metadata, you don't really need content."

Tempora, MUSCULAR → XkeyScore

Herman, Michael. Intelligence power in peace and war. Cambridge University Press, 1996.
Diffie, Whitfield, and Susan Landau. Privacy on the line: The politics of wiretapping and encryption. MIT press, 2010.
http://www.theguardian.com/world/interactive/2013/nov/01/snowden-nsa-files-surveillance-revelations-decoded

Richness of Metadata: Browser fingerprinting

- amiunique.org analyzes your browser's configuration (e.g., screen resolution, fonts, timezone, user agent) and compares it against a massive database of collected fingerprints to calculate how distinct your device is from everyone else’s

- It reveals which specific metadata points make your browser identifiable

- Proves that you can be tracked across the web even without using cookies or logging in

We need to protect the communication layer!
Why **anonymous communications?**

If you are a cyber-criminal!

DRM infringement, hacker, spammer, terrorist, etc.

But, also if you are:

Journalist

Whistleblower

Human rights activist

Business executive

Military/intelligence personnel

Abuse victims

Or you want to…

- Avoid tracking by advertising companies
- Protect sensitive personal information from businesses, like insurance companies, banks, etc.
- Express unpopular or controversial opinions
- Have a dual life

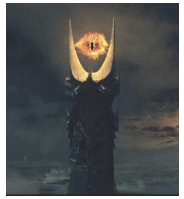
A professor who is a pro in LoL!

- Try uncommon things

…

https://www.eff.org/deeplinks/2013/10/online-anonymity-not-only-trolls-and-political-dissidents
http://geekfeminism.wikia.com/wiki/Who_is_harmed_by_a_%22Real_Names%22_policy%3F

Anonymous communications – Abstract model



IDs
Timing
Volume
Length
…

Bitwise unlinkability

Use cryptography to make inputs and outputs to the anonymous communication systems appearance (bits) different

(re)packetizing + (re)schedule

Destroy patterns (traffic analysis resistance)

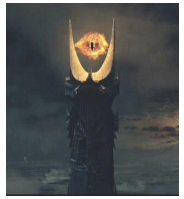
One-proxy problems

Low throughput

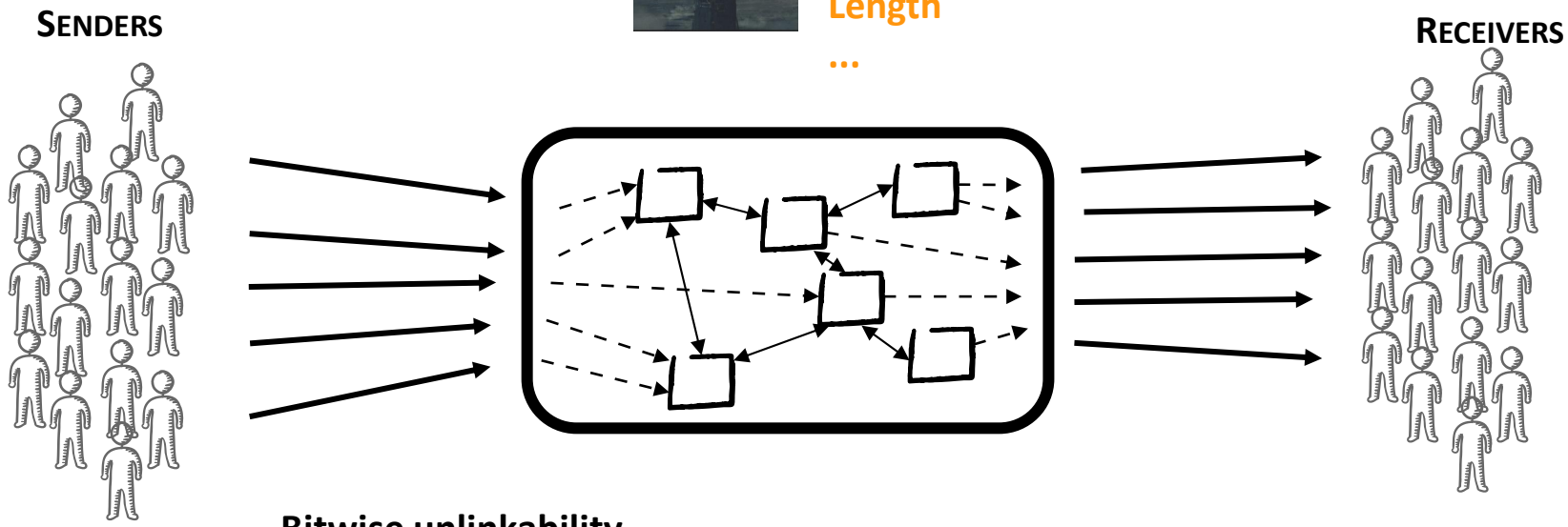
Corrupt Proxy or Proxy hacked / coerced

Real case: Penet.fi vs the church of scientology (1996)

Anonymous communications – Abstract model



IDs
Timing
Volume
Length
…



Bitwise unlinkability

Use cryptography to make inputs and outputs to the anonymous communication systems appearance (bits) different

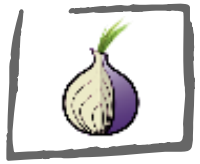
(re)packetizing + (re)schedule + (re)routing

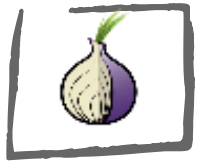
Destroy patterns (traffic analysis resistance)

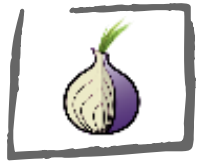
Load balancing

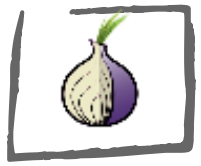
Distribute trust

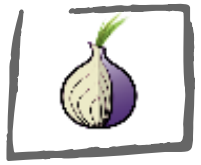
The Tor network – Onion routing

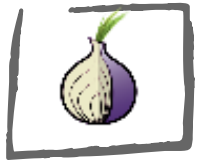




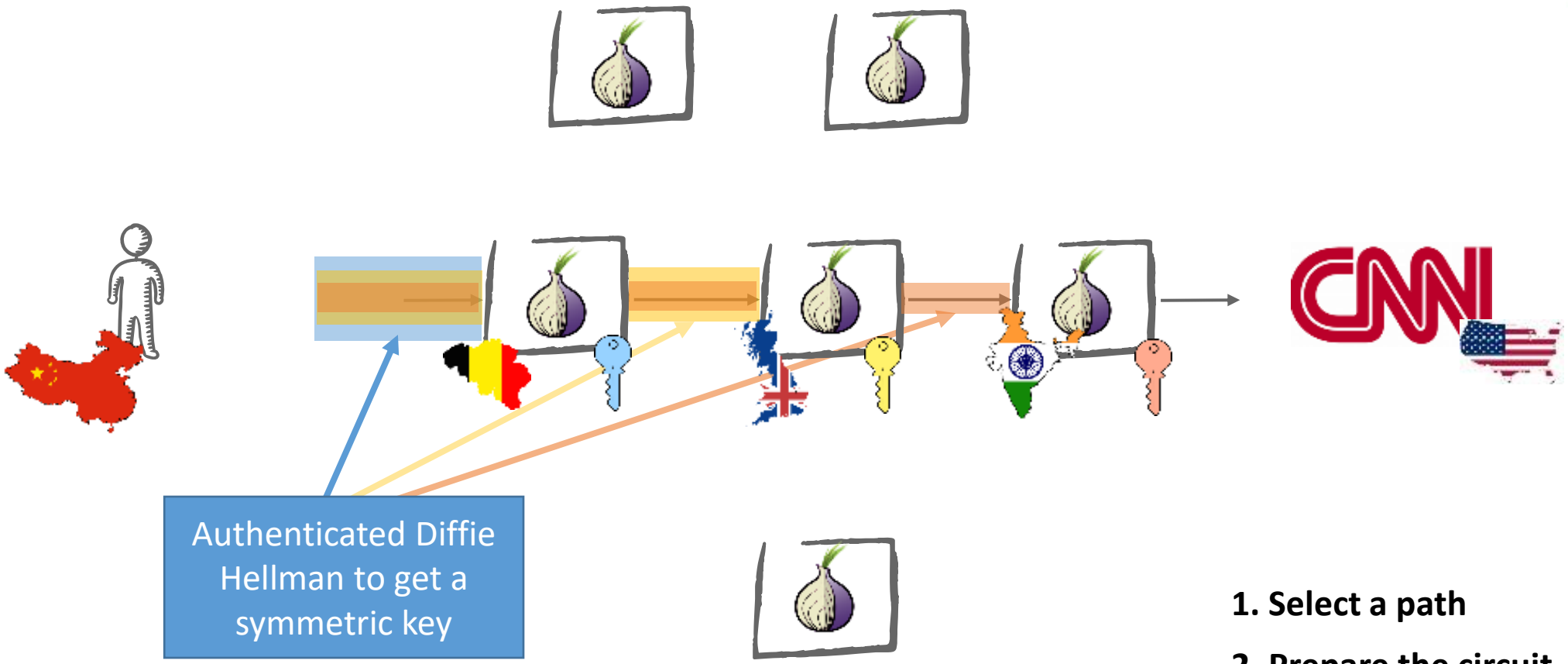








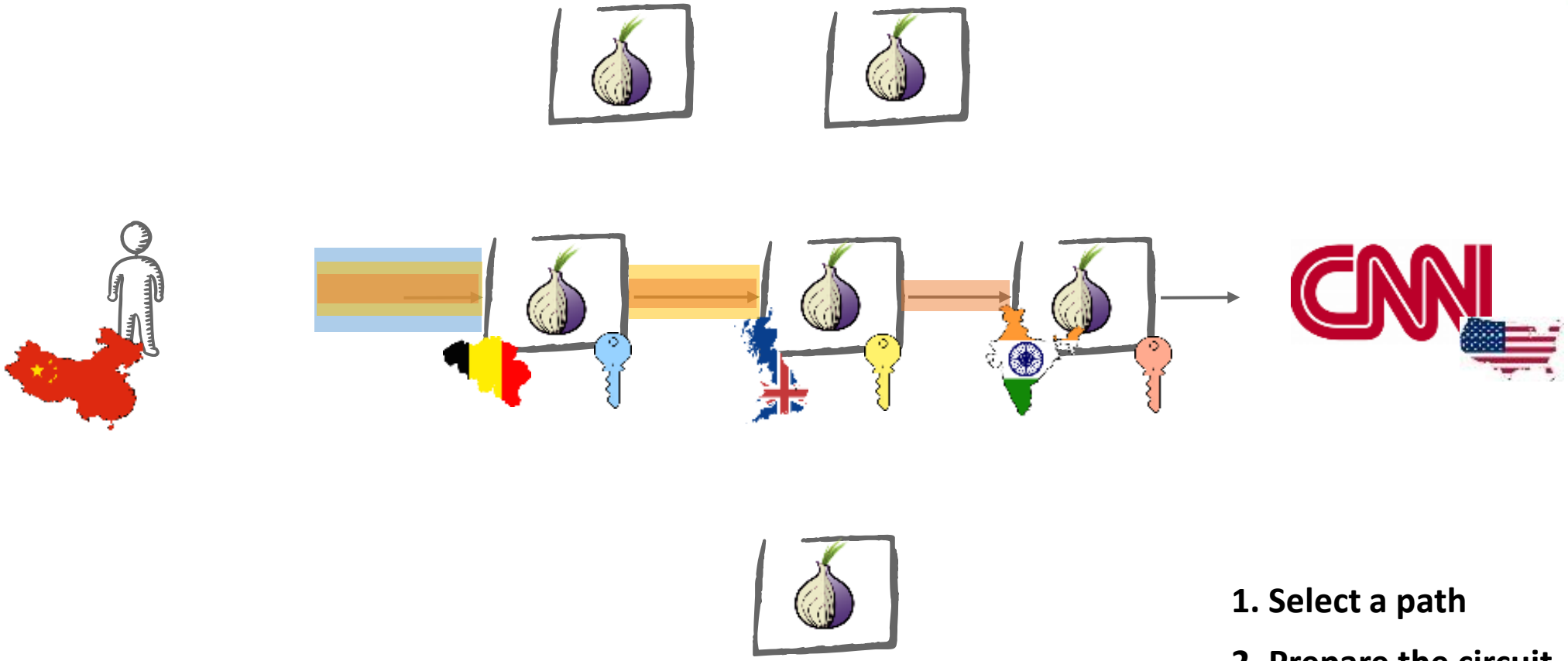
The Tor network – Onion routing



1. Select a path

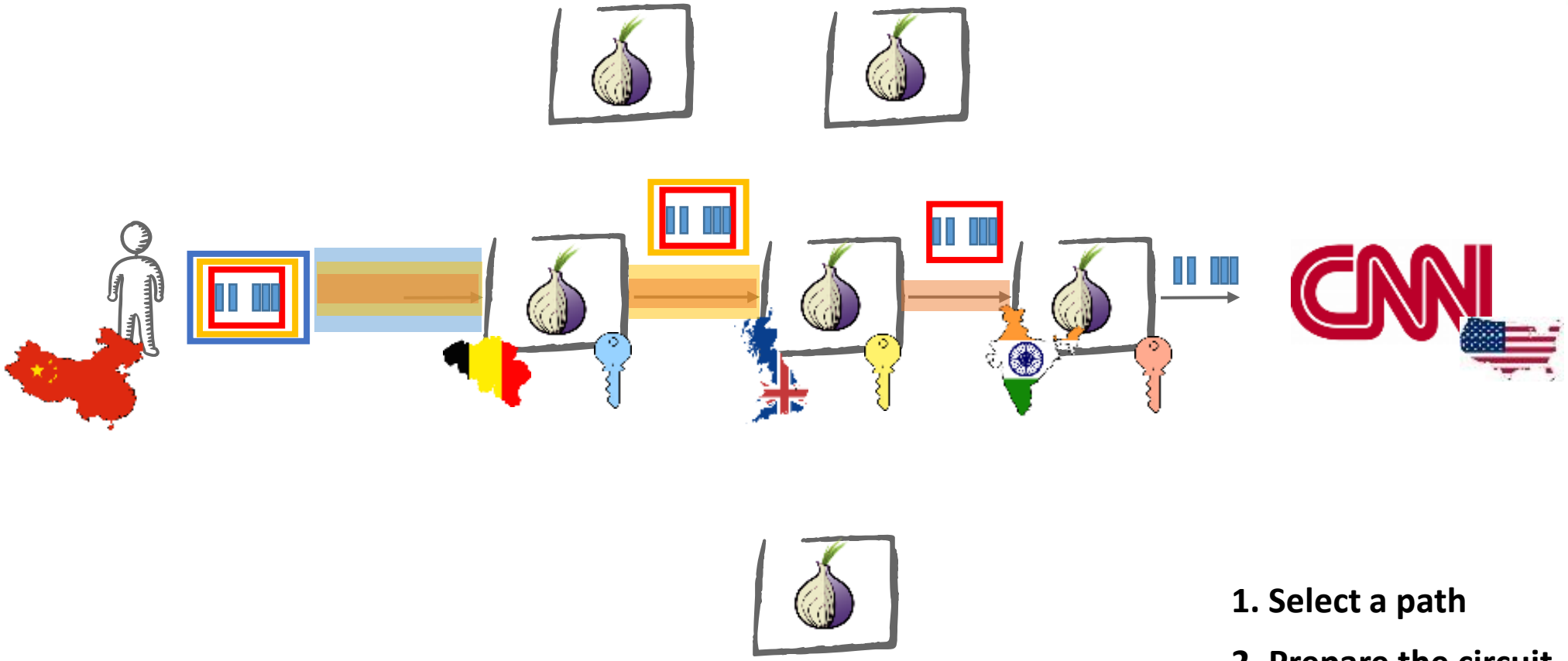
2. Prepare the circuit

The Tor network – Onion routing



1. **Select a path**
2. **Prepare the circuit**
3. **Send stream**

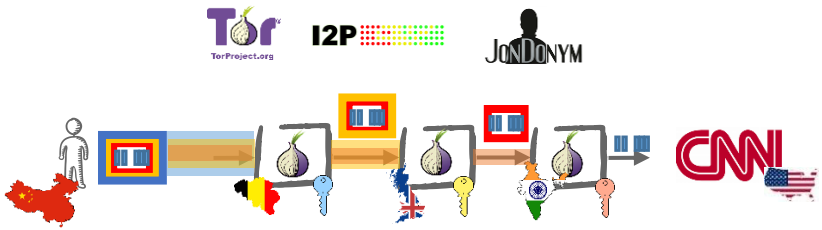
The Tor network – Onion routing



1. **Select a path**
2. **Prepare the circuit**
3. **Send stream**

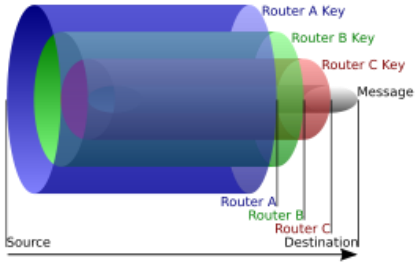
Anonymous communications out there

Low Latency

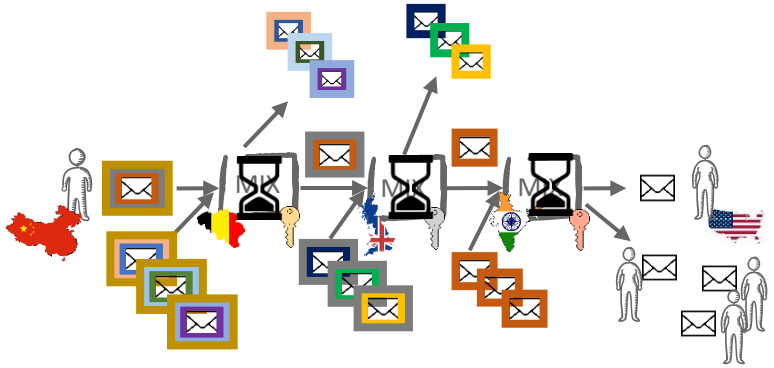


Web browsing, Instant Messaging, streaming

STREAM-based:  **fixed for the stream**



High Latency



Email, Voting,Bitcoin

MSG-based:  **vary every message**

One route per message + delays (slower!)

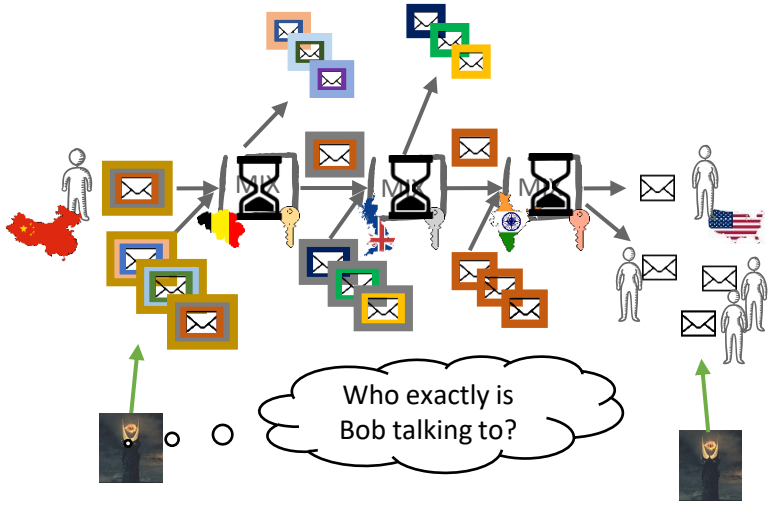
Anonymous communications out there

Low Latency

Volume of cells:
(e.g. how many cells per second)

Cannot resist **Global** Adversary
(Tor assumes that the adversary cannot see both edges)

High Latency



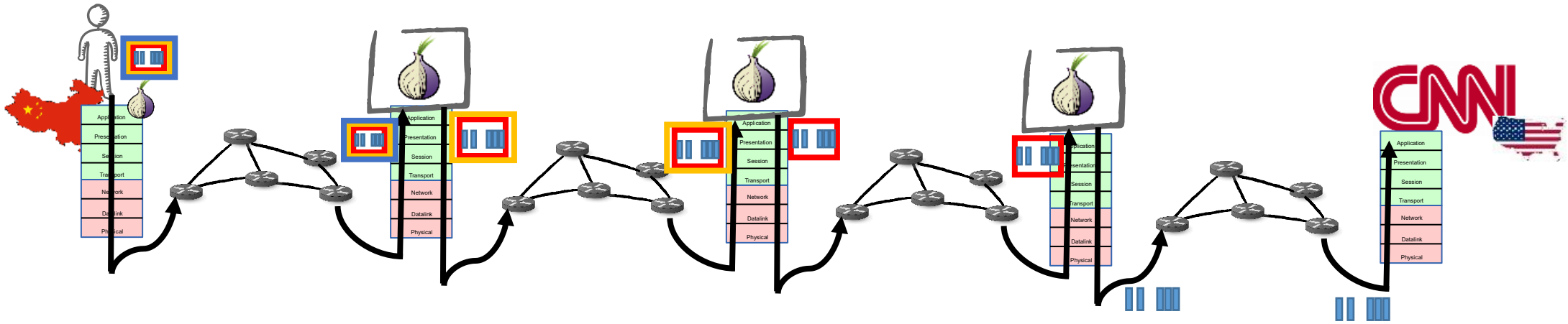
Global Adversary resistance
at the cost of latency
(and long term patterns revealed)

Anonymous communication networks are overlay networks

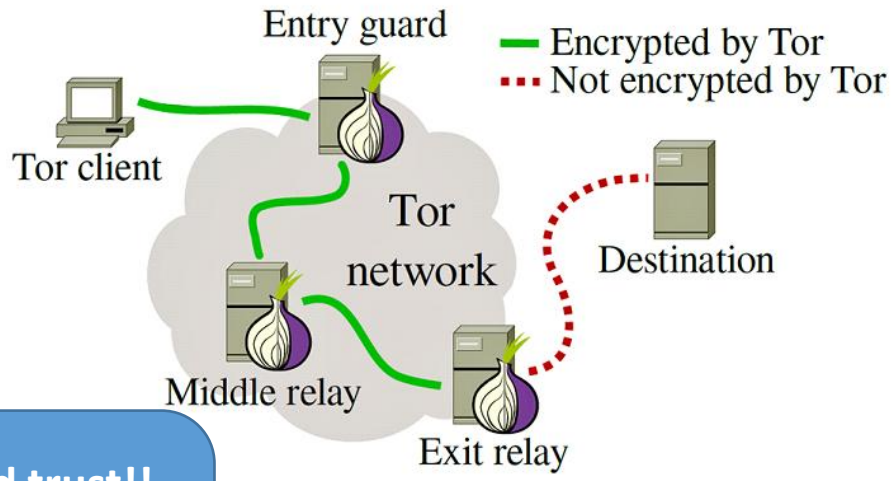
Nodes in anonymous communication networks (e.g., onion routers in Tor) are **not** internet routers. They work at the application layer!

(overlay network = a computer network that is built on top of another network)

A more realistic view of how Tor traffic travels would be this



Anonymous communications vs. VPN



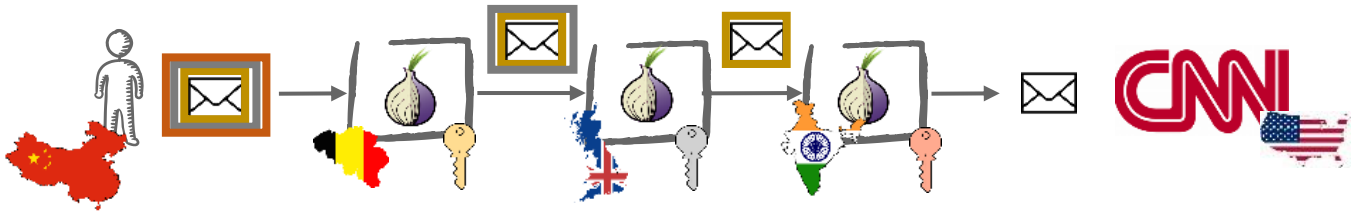
Different trust models!! Who is the adversary?

Decentralized trust!! Provides privacy as long as the adversary cannot see both edges

Centralized trust. No anonymity vs the VPN, or anyone seeing the VPN

https://www.privacyend.com/difference-between-tor-and-vpn/

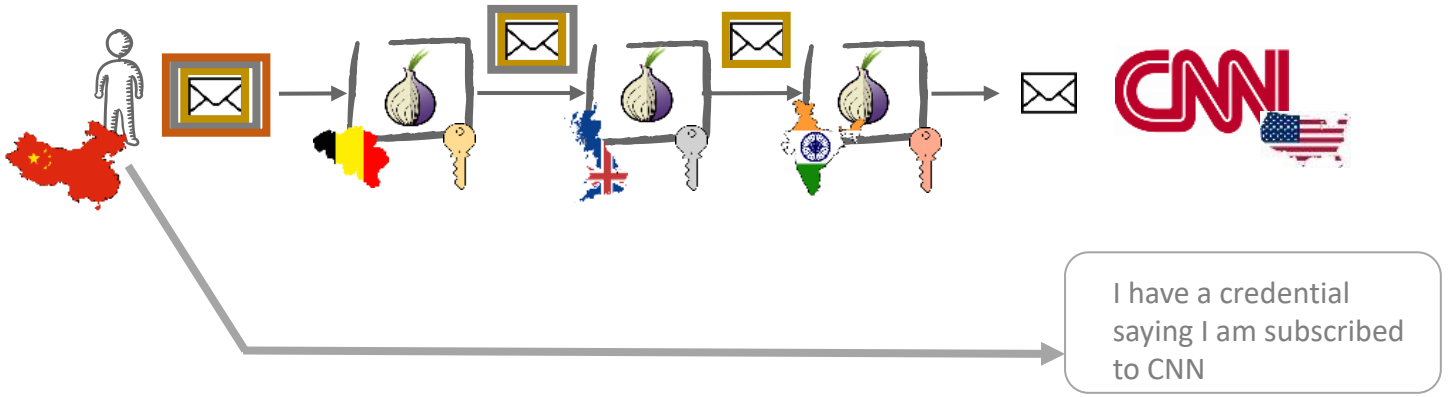
Anonymous communications at network layer what about the application layer?



me@cnn.com

me@cnn.com
logged in

Anonymous communications at network layer what about the application layer?



Anonymous credentials
Attribute-based credentials

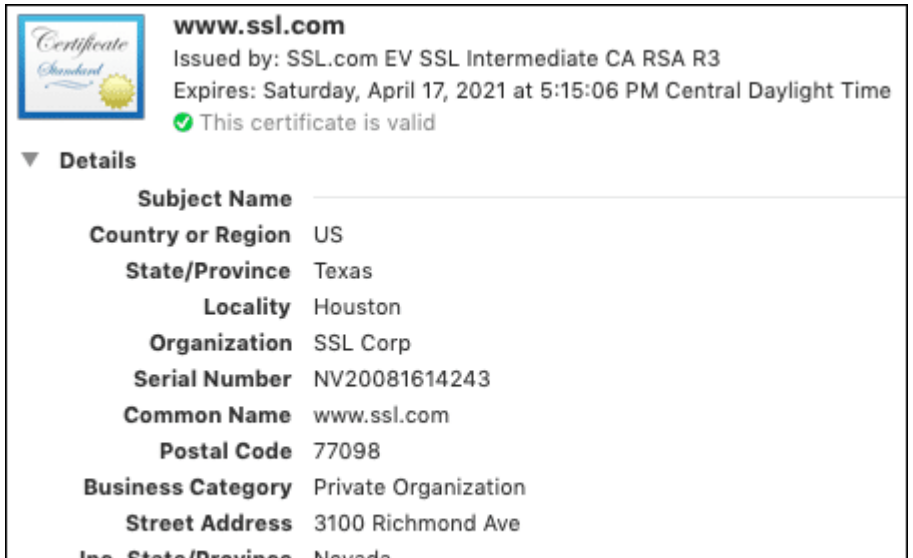
When shown the server **cannot**

- Identify Alice (if her name is not provided)
- Learn anything beyond the info she gives (and what can be inferred)
- Distinguish two users with the same attributes
- Link multiple uses of the same credentials

Public Key Infrastructure
(usual internet authentication)

Signed by a trusted issuer
Certification of attributes
Authentication (secret key)

No data minimization
Users are identifiable
Users can be tracked
(Signature linkable to other contexts where PK is used)



Attribute based credentials

Signed by a trusted issuer
Certification of attributes
Authentication (secret key)

Data minimization
Users are anonymous
Users are unlinkable across contexts

Other PETs examples

Private set intersection

a client and a server jointly compute the intersection of their private input sets in a manner that at the end the client learns the intersection and the server learns nothing (one-way PSI) or both learn the intersection (mutual PSI) -- private search

Blind Signatures

a server signs a message produced by a client without learning the content of the message -- eCash

Multiparty computation

parties to jointly compute a function over their inputs while keeping those inputs private –- compute total computations (statistics)

Private information retrieval

cryptographic method that allows a user to query a database without the server knowing which item the user asked for

Privacy Quantification

No Free Lunch Theorem [1]:

For every algorithm that outputs a D with even a sliver of utility, there is some adversary with a prior such that privacy is not guaranteed

[1] “No Free Lunch in Data Privacy”, Kifer and Machanavajjhala, 2011