

Privacy on the Network Level (Anonymous Communication Networks)

Iness BEN GUIRAT
iness.ben.guirat@ulb.be

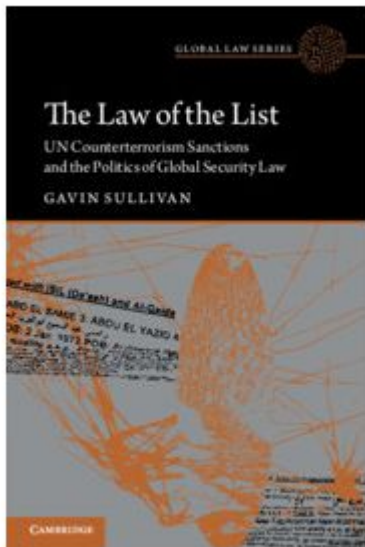
Who am I

- Ph.D from KU Leuven on Privacy with a focus on Anonymous Communication Networks
- Engineering degree in Networks and Telecommunications
- Currently a Postdoctoral researcher at Université Libre de Bruxelles
- Interested in the intersection of privacy and the socio-technical dimensions of technologies
 - Working with Prof. Jan Tobias Muehlberg and Prof. Jean-Michel Dricot

Outline

1. Privacy on the network level
 - a. Threat model
 - b. What is an Anonymous Communication Network (ACN)?
2. Tor
 - a. Overview
 - b. Attacks on Tor
3. Mixnets
 - a. Mixing strategies
 - b. Nym
4. Privacy Evaluation
 - a. Metrics

Why is Privacy important?



Protest surveillance today involves the acquisition, processing, generation, analysis, use, retention or storage of information about people engaging in protest just for participating in protests, without any regard to whether they are suspected of wrongdoing.

Protest surveillance negatively affects human rights, especially the right to privacy and freedoms of assembly and expression. Unjustified interferences with privacy prevents the enjoyment of other rights and they often provide the gateway to the violation of the rest of human rights freedom of movement, principle of non discrimination, as well as political participation.

THE PRESENT — FEBRUARY 20, 2019

How Spotify manipulates your

WIRED

BACKCHANNEL BUSINESS CULTURE GEAR IDEAS POLITICS SCIENCE SECURITY MERCH

SIDNEY FUSSELL

BUSINESS JUN 19, 2020 7:00 AM

How Surveillance Has Always Reinforced Racism

Sociologist and author Simone Browne connects the dots between modern marketing and the branding of slaves.

Search histories, location data, text messages.
How personal data could be used to enforce
anti-abortion laws



By Jennifer Korn and Clare Duffy, CNN Business

8 minute read · Updated 4:27 PM EDT, Fri June 24, 2022



VICE

Police Records Show Women Are Being Stalked With Apple AirTags Across the Country

Motherboard obtained reports of stalking, harassment, and abuse using AirTags, targeting victims of intimate partner violence.

Privacy is a security property!



Taken from "The Future of Security and Privacy", by Prof. Bart Preneel

Privacy is a security property!



A cyberattack on a government contractor exposed sensitive data of current and former Canadian government employees, including members of the Canadian Armed Forces and Royal Canadian Mounted Police personnel



From therecord.media

What is Privacy?

- *Intuitions* rather than exact properties?
- Security: Confidentiality, Integrity, Availability, Authentication
 - Guaranteed by Cryptographic schemes
- No cryptographic mechanism that guarantees privacy !

Privacy is a security property!

- SSH is used for secure remote login and file transfer. All data is **encrypted** and **authenticated**. What information can we extract about a password typed in a protected session?
 - **Note:** each key pressed is transmitted separately.
- Length of password is observable.
- Depending on the position of the key on the keyboard, different inter key timings.
- Attack (Song et al.): observe the inter key timings (many times) – infer what keys have been pressed.
 - Result: reduce the entropy of password – fewer guesses required.
- Note that there is still variability across different people. Adds noise – but also opportunities (Rubin et al.)!
- Monitor a user session and record the timings of key presses.
- Use existing profiles to infer their identities according to the leaked timing.
- Can extract both information and identity from a ‘secure’ session.

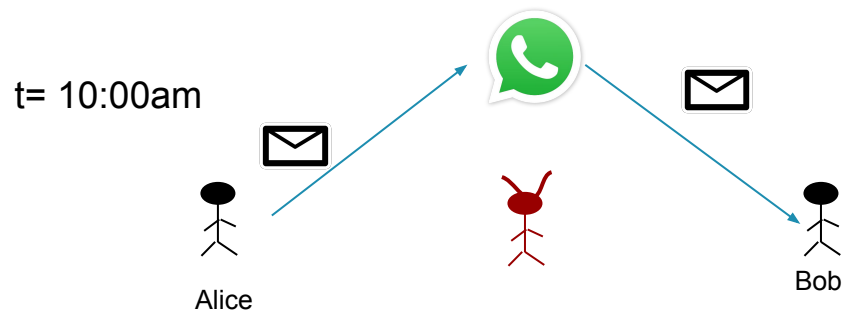
Privacy on the network level

- Was the Internet supposed to be private/anonymous?
- Web 1.0: Static web pages
- Web2.0: Social media accounts
- Privacy violations by Facebook, Google etc...
- Privacy on social networks
- but even then:
 - 5.35 billion internet users
 - ~200 people using the Cyberus Guest Wifi
 - A dozen have downloaded a particular paper
 - not even mentioning website fingerprinting and other PII
- Internet Protocols such as TCP, UDP, HTTPS were not designed with Privacy in mind (leak metadata by default)

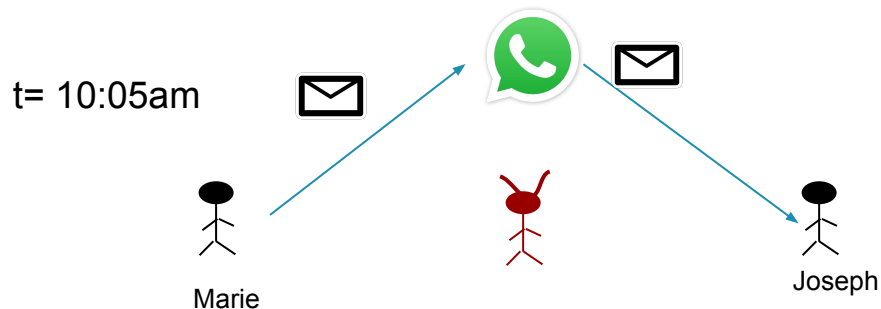
How easy to collect metadata

- Exposed by default in core internet protocols: TCP/IP, HTTP, UDP, FTP, TLS, DNS, ...
- Available to a large number of intermediaries
 - Local LAN or WiFi router
 - Internet Service Provider (ISP), Mobile network operator
- Metadata has lower legal protection than data content
- Metadata is machine-readable, lower volume than content and much easier to interpret automatically than content
- Metadata is difficult and expensive to protect

How easy to collect metadata



Privacy on the network level:
Anonymous Communication
Networks (ACNs)



The adversary in Anonymous Communication Networks (ACNs)

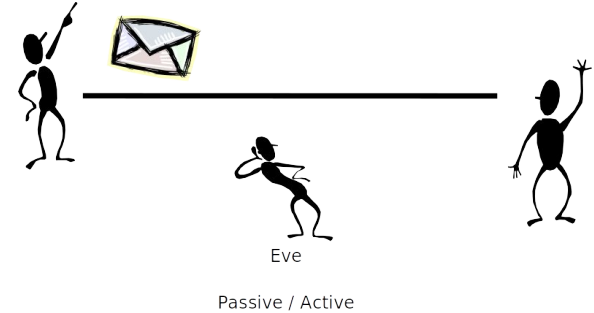
- Adversary can't / won't break cryptographic primitives: **content is encrypted**
- Collects **metadata: Traffic Analysis Techniques**
- Goal: de-anonymize subjects and extract information.

What network metadata?

- Identities of communicating parties.
- Time, duration or length of transmissions.
- Location of sender or receiver.
- No content – encrypted.

Why is traffic analysis **so valuable** as opposed to 'cryptanalysis'?

- It provides lower quality information compared with cryptanalysis, but it is both easier and cheaper to extract and process.



“Traffic analysis, not the cryptanalysis, is the backbone of communication intelligence”
Whitfield Diffie & Susan Landau

Is metadata important?

- Abortion clinic: “I Know Why You Went to the Clinic” by Brad Miller, Ling Huang, A. D. Joseph, and J. D. Tygar
- “Unique in the shopping mall: On the reidentifiability of credit card metadata” by Yves-Alexandre De Montjoye, Laura Radaelli, Vivek K. Singh, Alex ‘Sandy’ Pentland
- NSA General Counsel Stewart Baker: “Metadata absolutely tells you everything about somebody’s life. If you have enough metadata, you don’t really need content.”

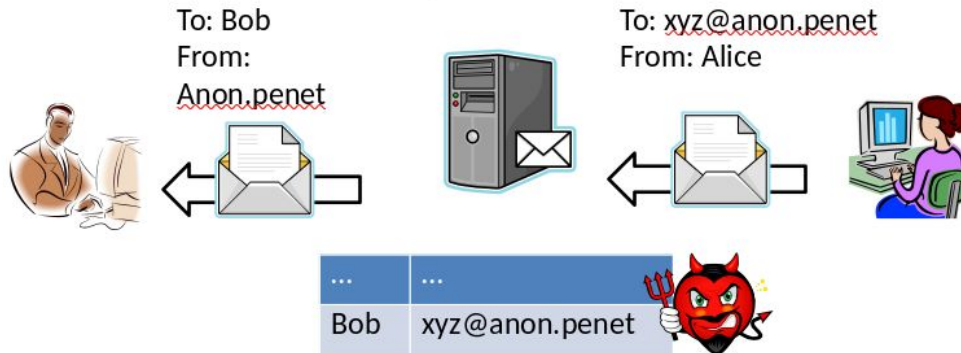
Anon.penet.fi: pseudonymous remailer (1993)

Should people be required to tie their real name to their online communications?

Simple proxy, substituted email headers

Kept table of correspondences nym-email

Supported anonymous replies



Anon.penet.fi: Attacks?

Brought down by “legal attack” in 1996

Lesson learned: do not keep tables of correspondences!

Protection of users, but also protection of services themselves

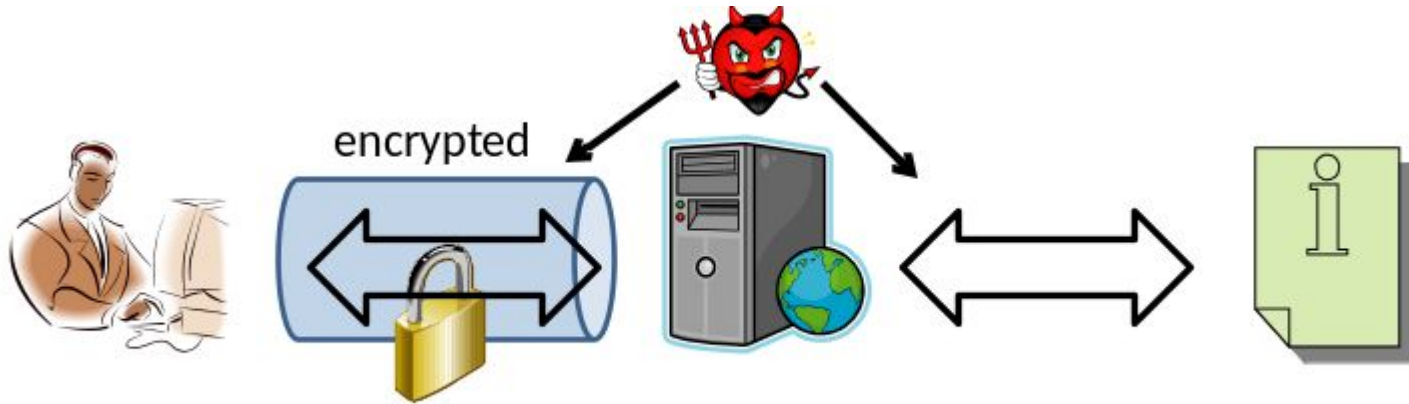
Trivial to find correspondence by observing the server

Anonymizer and SafeWeb (mid-90s)

Web proxies: strip identifying information and forward

Main difference from anon.fenet?

Keeping long-term logs is not needed (communication always initiated by the user)



Anonymizer and SafeWeb: Attacks

Vulnerable to attacks that correlate traffic to and from the server

The anonymity provided depends critically on the integrity of the company operating the service and of its staff !



Privacy and Security Fanatic

Ms. Smith

[◀ Previous Post](#)

[Next Post ▶](#)

Anonymizer tied to company selling TrapWire surveillance to governments

Anonymizer, which is known for selling "powerful online privacy and security" services, has ties to Abraxas Corporation and the TrapWire spying system being sold to governments on a global scale.

By [Ms. Smith](#) on Tue, 08/14/12 - 4:53pm.

What would a “perfectly private” communication network offer?

The possibility for Alice to communicate while preventing adversaries from learning:

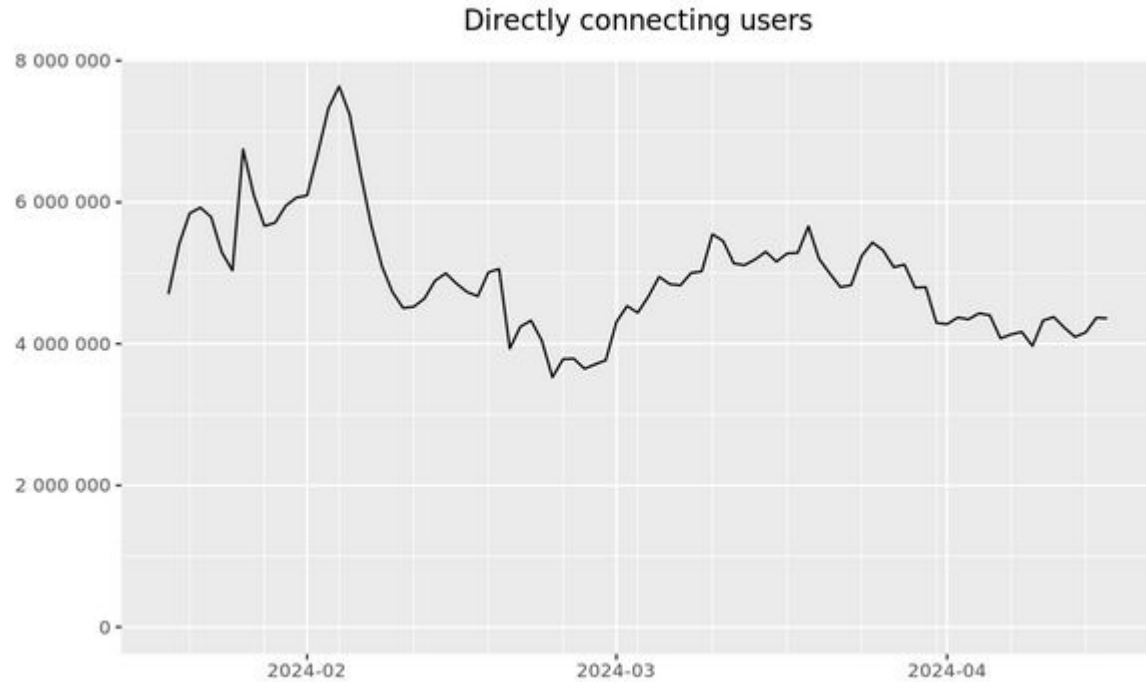
- **What, When, From where, How long**
- **With whom**
- The **amount** of data she is sending or receiving
- Any **patterns** in her communications
- **Whether** she is communicating at all

Tor: History of development



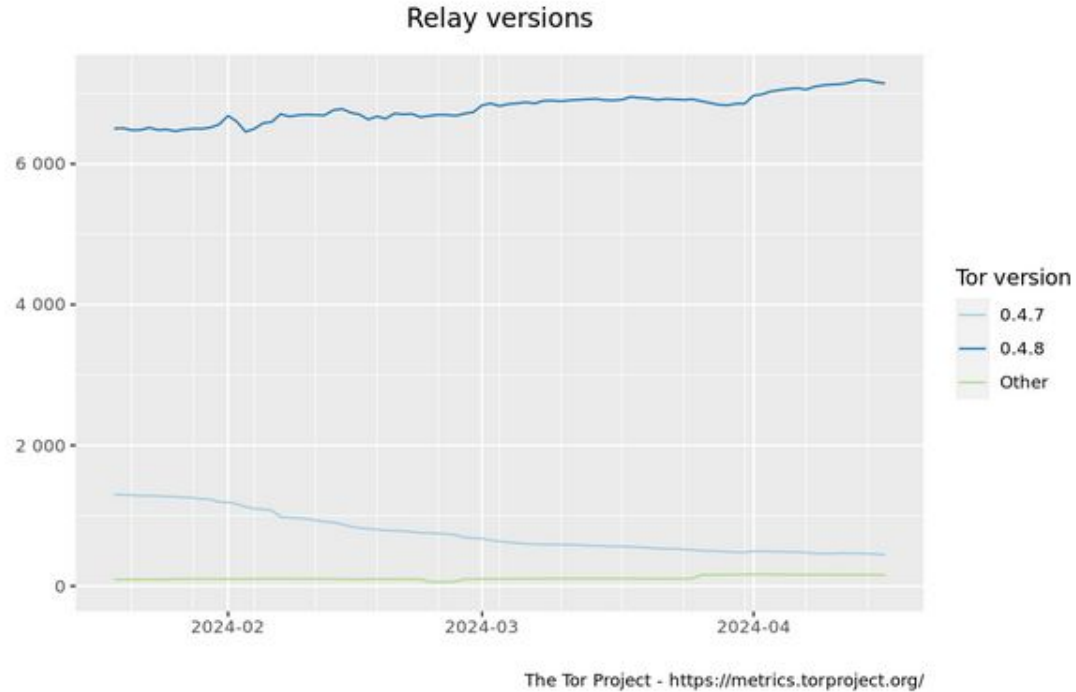
- Developed at the US Navy Research Lab (1996)
 - Primary purpose: protecting government communications
 - Need to “mix” with civilians!
- ZK Freedom Network.
 - Canadian company, commercial project (1999-2001)
 - Failed
- Second-generation Onion Routing: Tor (since 2003)

Tor: Users

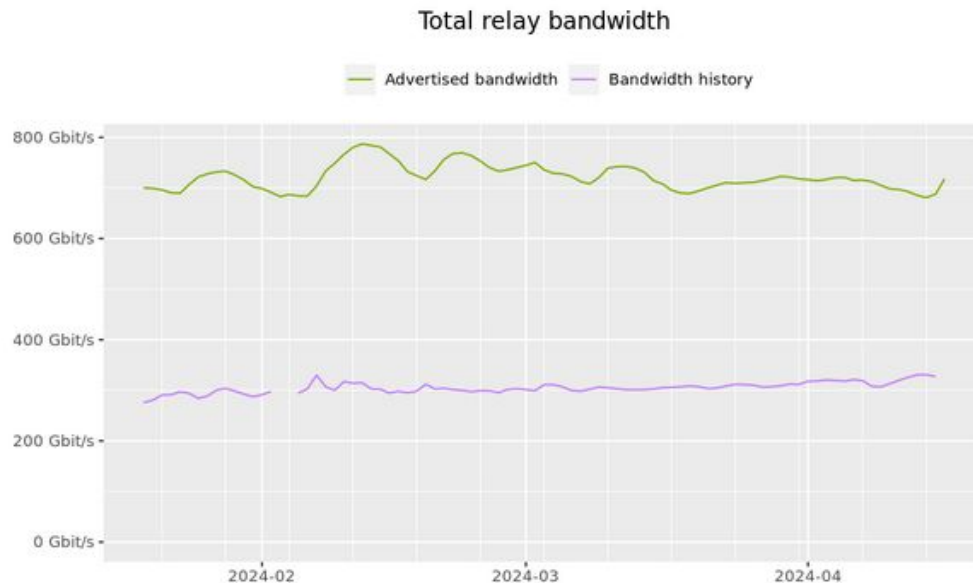


The Tor Project - <https://metrics.torproject.org/>

Tor: Relays



Tor: Bandwidth

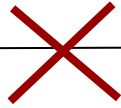


The Tor Project - <https://metrics.torproject.org/>

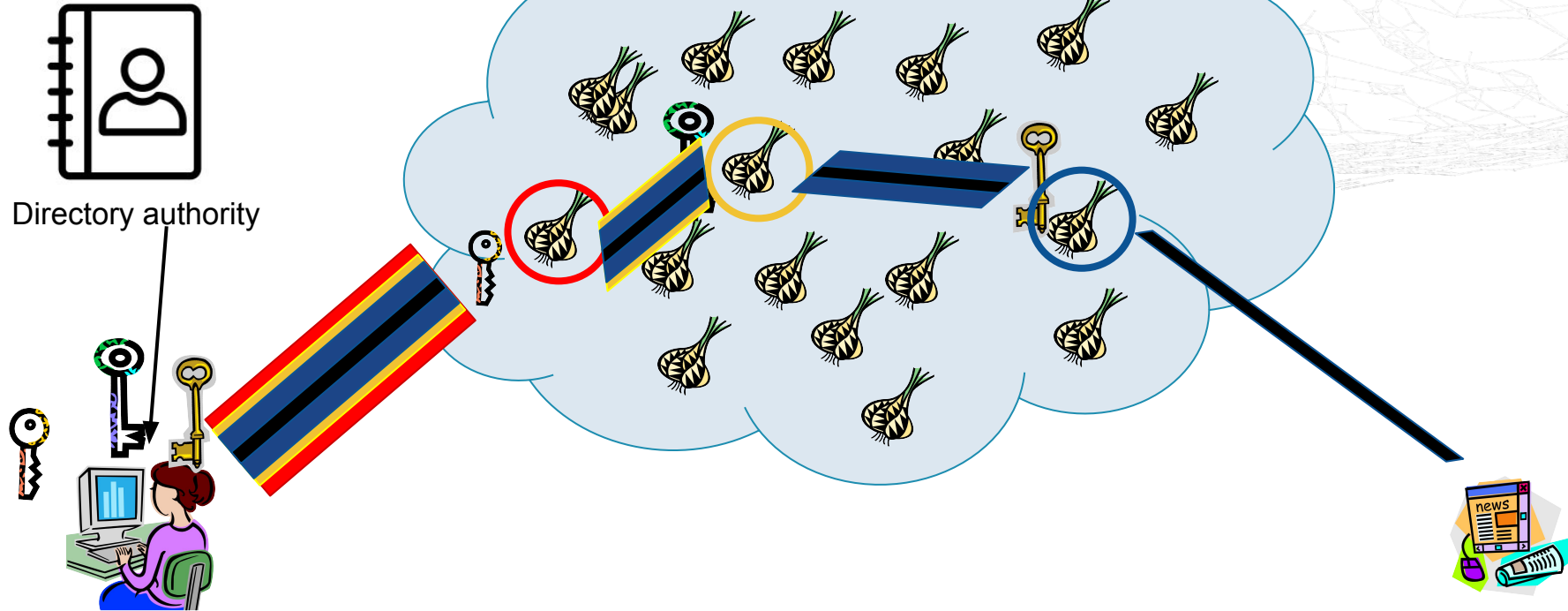
Advertised bandwidth: the volume of traffic, both incoming and outgoing, that a relay is willing to sustain, as configured by the operator and claimed to be observed from recent data transfers.

Bandwidth history: the volume of incoming and/or outgoing traffic that a relay claims to have handled on behalf of clients

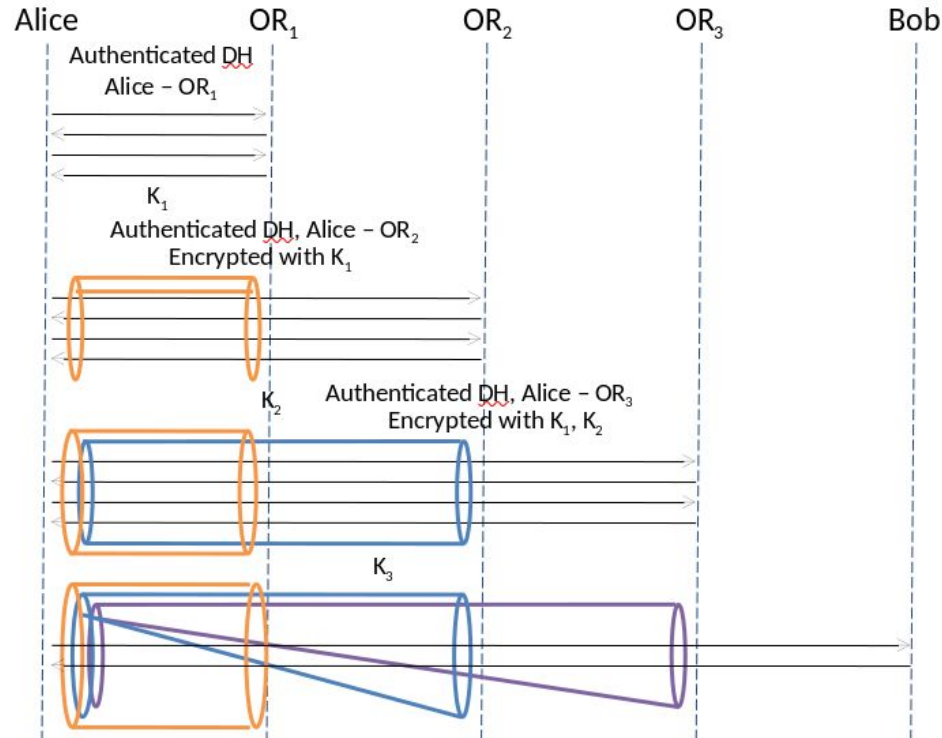
Tor: Network



Tor: Network



Tor: Circuit-Based Communication



Tor: Relay Keys

- long-term "Identity key"
 - signing-only
 - establish relay identity
- medium-term "Onion key"
 - rotated once a week
 - used to decrypt onion skins when accepting circuit extend requests
- short-term (ephemeral) "Connection key" used to negotiate TLS connections
 - discarded when the circuit ends
 - "connection-based"

Tor: Attacks



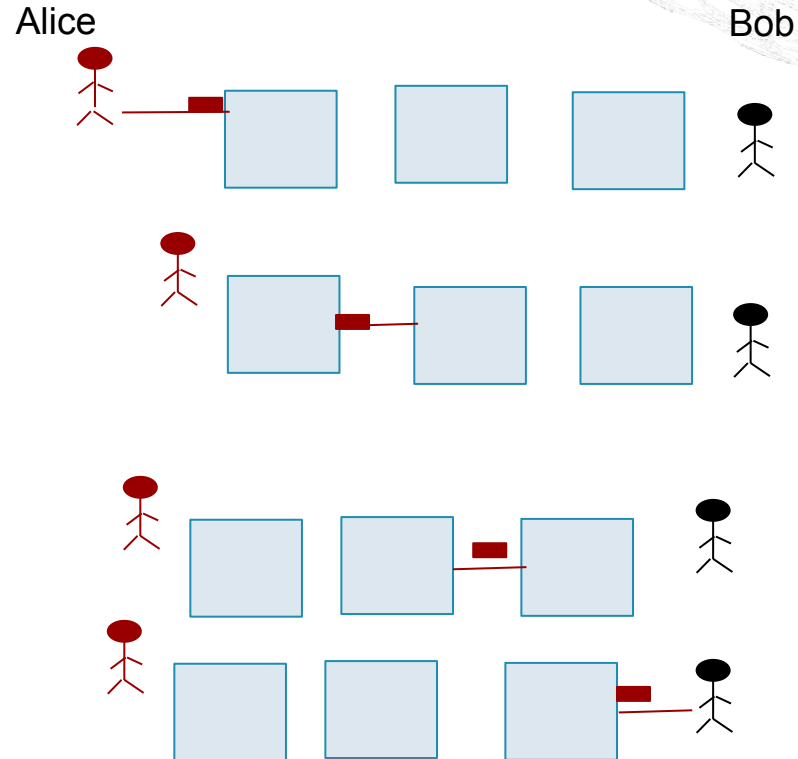
The directory authorities (~10) provide a signed list of all the known relays: specifying their onion keys, locations, exit policies, etc.

Question: Can you trick users into connecting to your (adversary) nodes instead of the Tor nodes? What would you need to do?

Location and public keys of the directory authorities **hard coded** into the Tor software

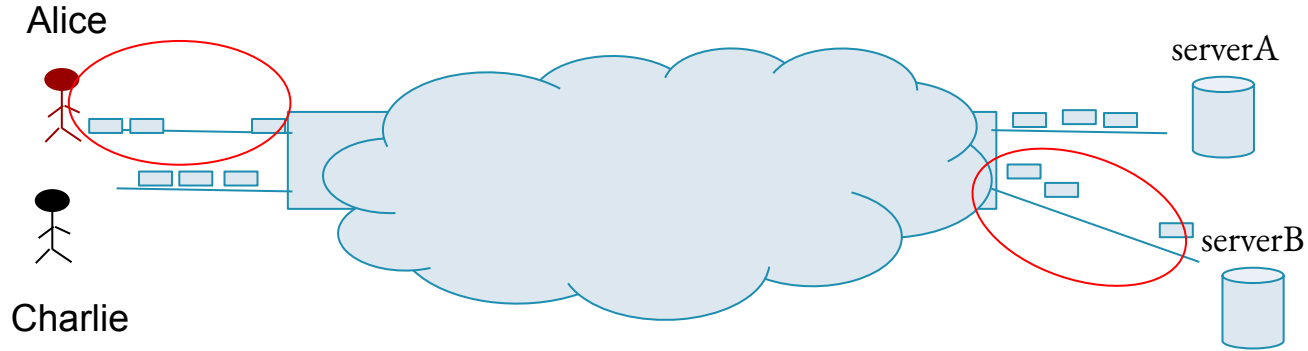
Tor: Attacks

How: Adversary observes **all** inputs and outputs



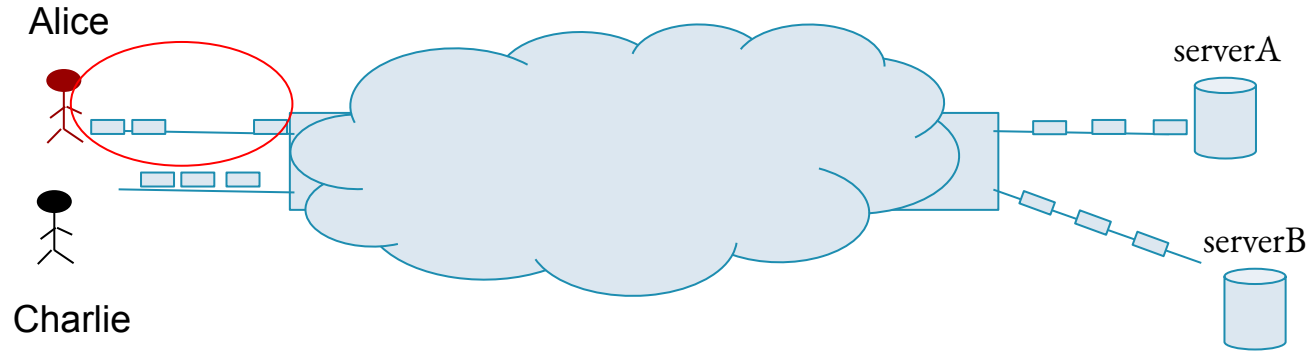
Tor: Attacks

How: Adversary observes **patterns** of inputs and outputs



We need better !!

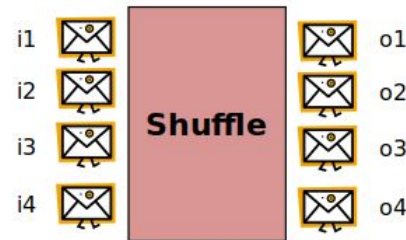
Add delays



Mixnets

Mix: Proxy for anonymous email

Goal: an adversary observing the input and output of the mix is not able to relate input messages to output messages



Bitwise unlinkability

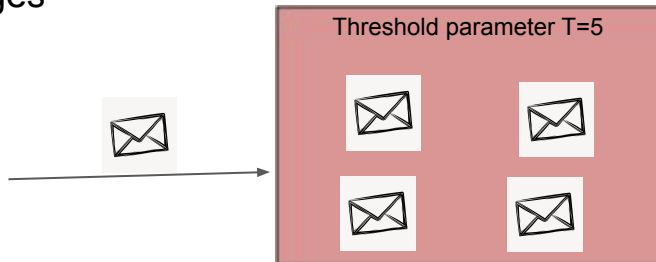
Input messages can not be correlated to **output** messages based on content or size ==> cryptographic packet format such as sphinx

Prevent traffic analysis based on message I/O order and timing

Achieved by batching and shuffling messages => Various mixing strategies

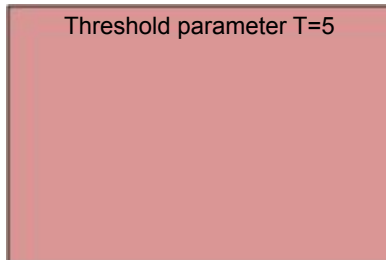
Mixing strategies: Threshold

Phase1: Collect messages



Advantage: Guarantees anonymity
At the cost of:
Latency depends on traffic

Phase2: Mix and Flush



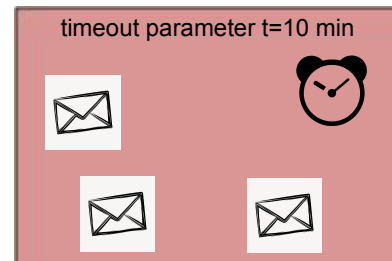
Mixing strategies: Timed

Timed: waits until a timeout then shuffle

Advantage: Bound end-to-end latency

Cost of:

Anonymity depends on traffic



Mixing strategies: Poisson

- Memoryless property of exponential distributions
- Delay each message individually with the amount of time drawn from an exponential distribution
- Delays picked by the sender: can predict delivery time
- Cost of:
 - Some messages never leave

For an exponential random variable X it holds that:

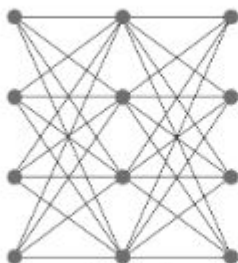
$$\Pr[X > a+b \mid X > a] = \Pr[X > b]$$

Mixnet Topologies: Distribute Trust

- Roles: distinguish participants (endpoints, mixes)?
- Length of packet paths: number of hops/layers \Rightarrow anonymity
- Anonymity: adversarial strategy
- Performance: average mix compute capabilities, path length,
- Scalability: traffic volume and frequency known?
- Availability: alternative paths if mixes fail
- Operational: key distribution, availability/churn, networking



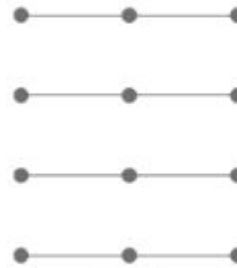
Free Route



Stratified



Stratified Restricted



Cascade

The adversary in mixnets: Passive

Objectives:

- **Identify** the sender (receiver) of a received (sent) message
- **Link** messages as sent (received) by the same user
- **Profile** the activity of users (time and volume of sending/receiving, likely recipients)

View on the network:

- **Global**: can observe all communication links and look at all traffic (GPA)
- **Partial**: observes some communication links and traffic (adaptive or non-adaptive)
- **Local**: controls one edge of the network (Alice's ISP, employer, or malicious sender/receiver)

Passive adversary who in addition can:

- **Inject** messages at any point of the network (not just users)
- **Delete** or **delay** messages, for DoS and traffic analysis
- **Modify** messages to help with tracing

The adversary in mixnets: Active

Corrupt Insiders:

- Some nodes in the anonymous communication infrastructure belong to the adversary
- Or many, if the system is vulnerable to **Sybil attacks**
- They leak all the secrets they know and coordinate (**correlation** between inputs and outputs)

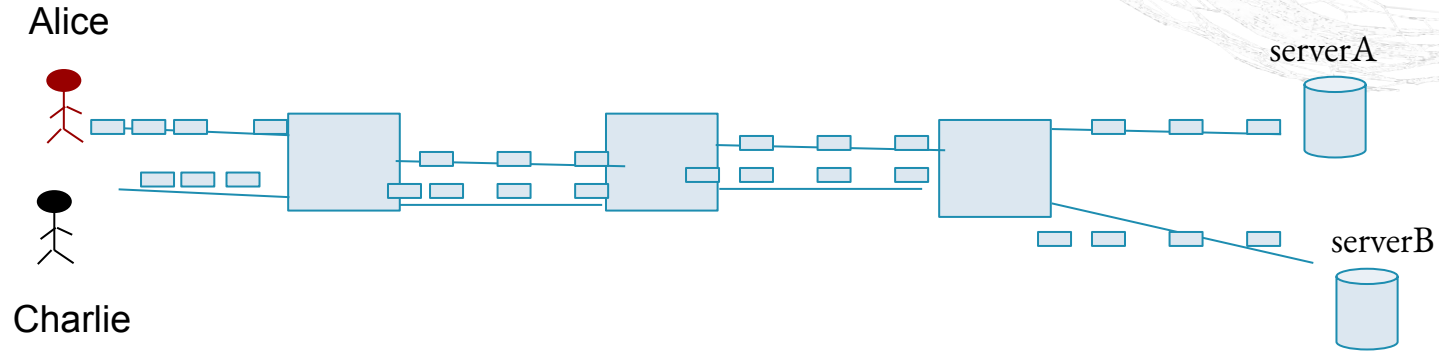
Coercion

- Honest nodes may be forced to **cooperate** with the adversary through blackmail, bribery, legal or physical threats

They should:

- Know as few secrets as possible
- Have the option to (plausibly) lie

Mixnets: Attacks



Counting packets

Mixnets: Long-term intersection

Method:

Combine many observations (looking at who receives when Alice sends)

Intuition:

If we observe rounds in which Alice sends, her likely recipients will appear frequently

Result:

We can create a vector that expresses **Alice's sending profile**

Hard to conceal persistent communications (also in low-latency systems!)

Dummy traffic can help but expensive

Mixnets: (N-1) Blending attack

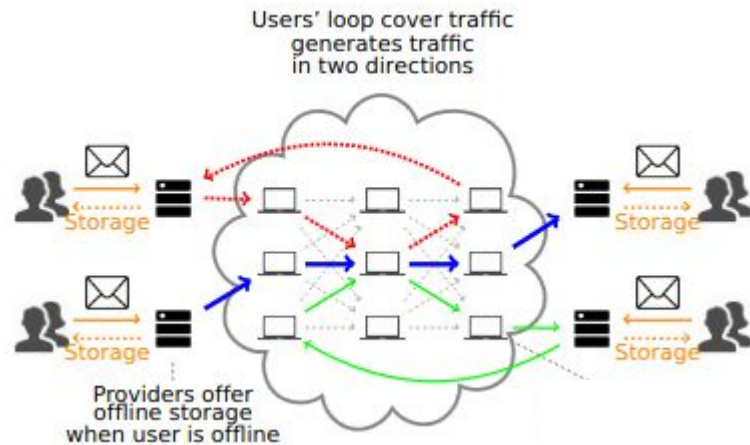
1. Empty the mix from legitimate messages
2. Let the target message into the mix
3. Fill the mix with attacker-generated messages, while preventing other legitimate messages from entering the mix
4. At the time of flushing the adversary recognizes his own messages. The unknown message is the target

Can be mitigated with:

1. dummy traffic
2. Verifiable shuffle

Nym Technologies

- Based on the paper “The Loopix Anonymity System”
- Poisson mixes
- Layered Topology: Scalable!
- Gateways for access control (payments for usage) and receiving messages by offline clients



Nym Technologies

Each client has a service provider:

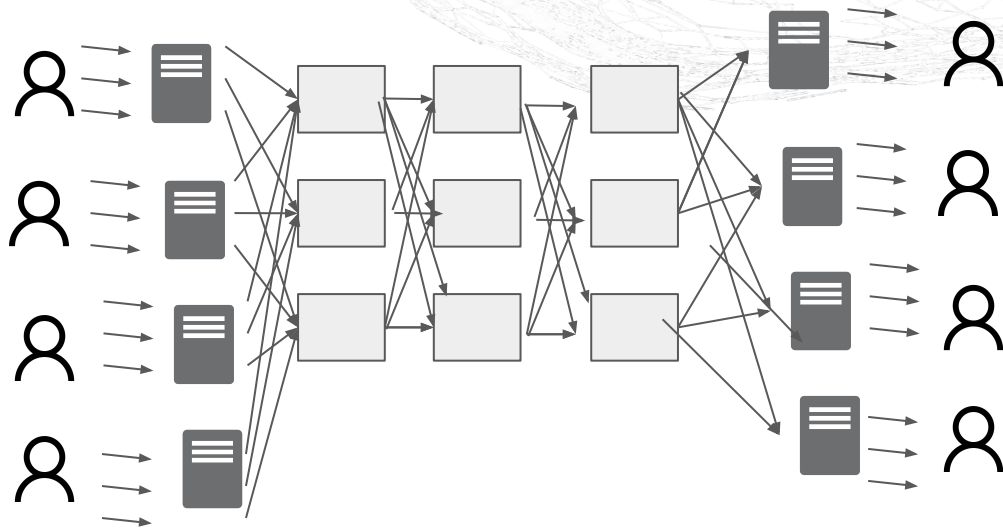
- Entry and exit for the mixnetwork
- Store message for offline clients
- Can be used for access control (for example payment)

Each stream of traffic follow a poisson process: if no message then substitute by a dummy message

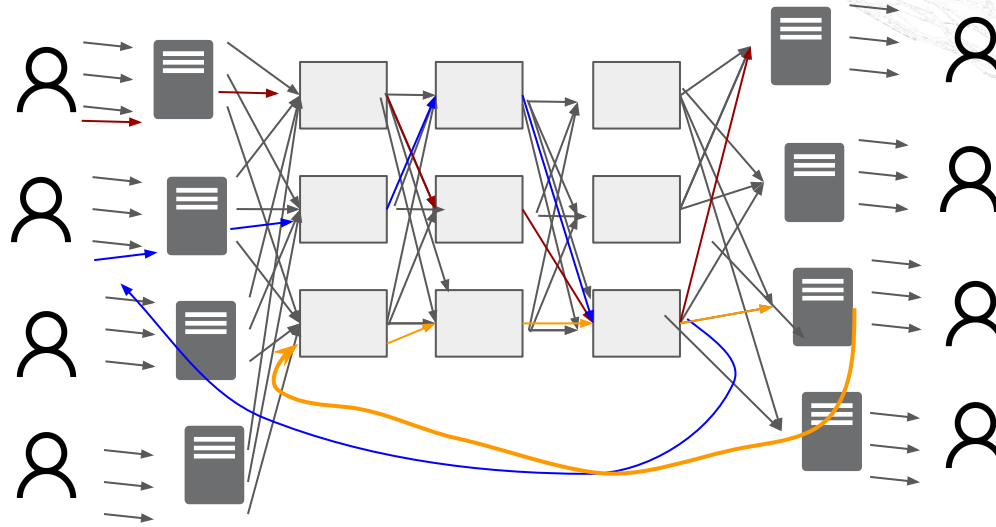
Mixing strategy: Poisson mixing

Packet based source routed: why not circuit?

at the cost of?



Nym Technologies: Dummy Traffic



1. Drop cover traffic: dropped by the provider (dropped according to a flag)
2. Client loop messages: start and finish at the same entity: detect N-1 attack
3. Loop generated by the mixes

Nym Technologies: Adversary

1. GPA:

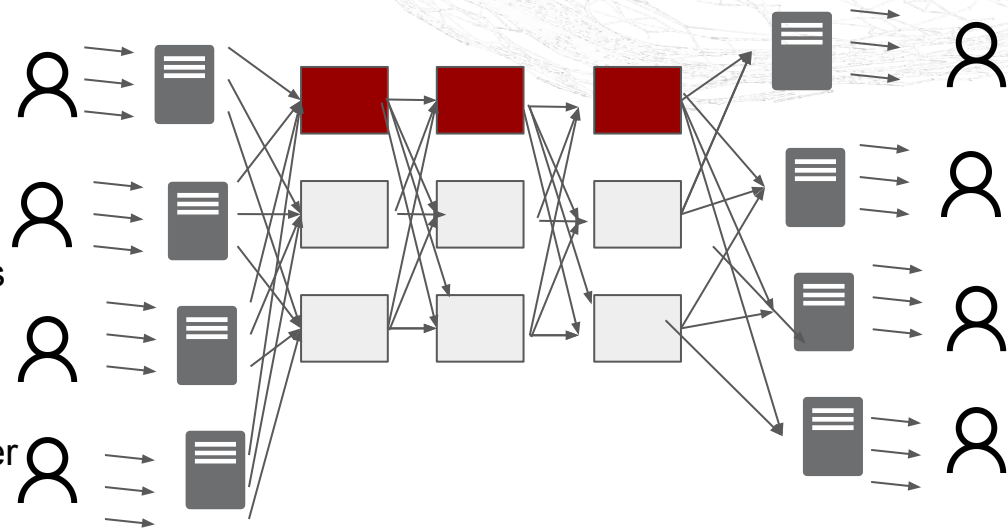
Increase delays increase anonymity
Increase number of layers
Anonymity Trade-off

2. Active Adversary: able to compromise mixes

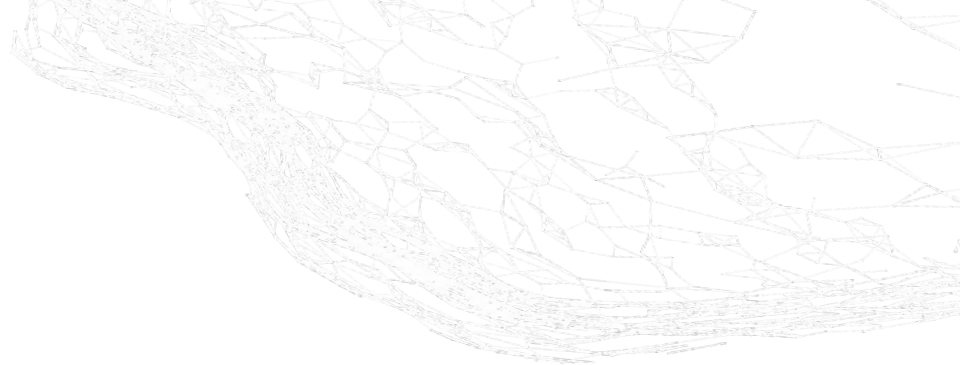
Key distribution: Validators maintain
blockchain with network info

As long as there's one honest node per
path

Epoch: change positions: problems?



Mixnet vs Tor



Similar

- Source routed with nested encryption
- Packets traverse a network with multiple hops

Different:

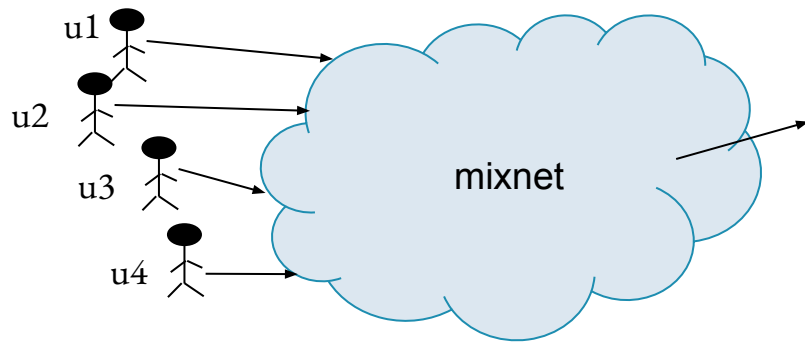
- Tor is connection-based vs Mixnets that are packet-based
- Tor does not add latency vs latency added in Mixnets
- Vulnerable to end-to-end confirmation vs (possibly) vulnerable to long-term intersection attacks
- Designed to resist local adversaries vs global adversaries
- Additionally (possible in both systems but easier to design for mixnets): Dummy traffic strategies to strengthen anonymity and enable unobservability

Privacy Metrics

Anonymity: “the state of being not identifiable within a **set** of subjects, **the anonymity set**”

Number of subjects in the anonymity set

- Given a target member $u1$, it is defined as the (size of the) set of members the adversary cannot distinguish from $u1$
- The larger the anonymity set, the more anonymity a member is enjoying
- Simplicity
- Only depends on the number of members in the system

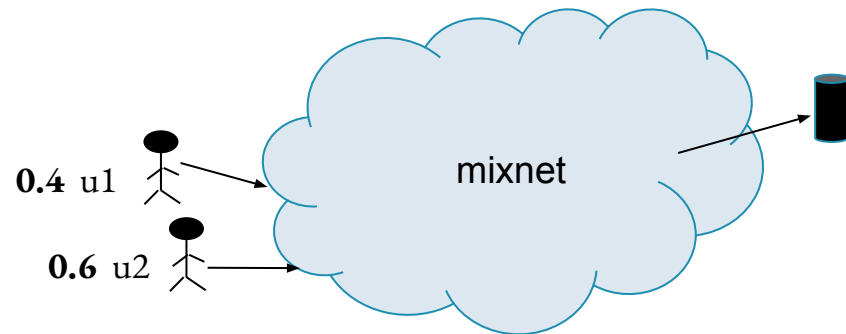
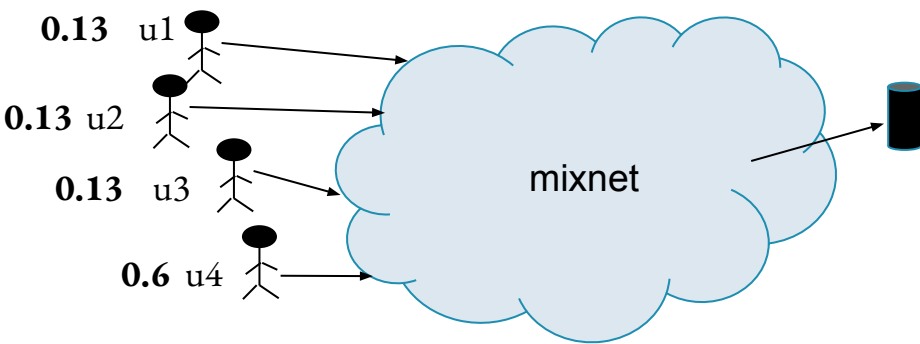


What if **not all of them** appear to be the target with equal likelihood?

Probability assigned to a subject:

worst case: user with highest probability is chosen as sender/receiver

Privacy Metrics



Anonymity depends on both:

1. The number of subjects in the anonymity set
2. The probability of each subject in the anonymity set being the target

Privacy Metrics: Entropy

System 1	System 2	System 3
{0.1,0.1,0.1,0.7}	{0.25,0.25,0.25,0.25}	{0.2,0.3,0.4,0.1}

$$H(X) = - \sum_{i=1}^N p_i \log_2(p_i)$$

The entropy metric defines the degree of success of the attacker // describes the uncertainty of the attacker

If $H = 0$, the attacker has succeeded with 100% confidence in tracing the target message.