Gradiant

# Bayesian inference to evaluate information leakage in complex scenarios

Carmela Troncoso
Gradiant, Spain
12th June 2014

GALICIAN RESEARCH AND DEVELOPMENT CENTER IN ADVANCED TELECOMMUNICATIONS

# Privacy beyond encryption

- Common belief: “if I encrypt my data, then the data is private”
  - Encryption works and gets more and more efficient!
  - But does not hide all data
    - Origin and destination
    - Timing
    - Frequency
    - Location
    - …

- These data contain a lot of information
  - WWII: The English recognized German Morse code operators
  - Nowadays:
    - *Phonotactic Reconstruction of Encrypted VoIP conversations: Hookt on fon-iks*. A. White, A. Matthews, K. Snow, and F. Monrose. S&P11.
    - Peek-a-Boo, I Still See You: Why Efficient Traffic Analysis Countermeasures Fail. Dyer, K. P., Coull, S. E., Ristenpart, T., & Shrimpton, T. S&P12
    - I Know Why You Went to the Clinic: Risks and Realization of HTTPS Traffic Analysis.
      *Brad Miller, Ling Huang, A. D. Joseph and J. D. Tygar. PETS 2014*

# Easy, let's hide this information!

- Delay messages to change frequency and timing patters
  - Messages cannot be delayed for too long

- Add dummy events to confuse the adversary
- Pad packets to hide their length
  - Bandwith is in general limited

- Reroute messages to hide origin and destination
  - Delays messages
  - Needs of collaboration or dedicated infrastructure

- Obfuscate the location
  - Obfuscation must not prevent usability

# Maybe is not that easy…

- Design decisions to:
  - Balance available resources and privacy
  - Balance usability and privacy

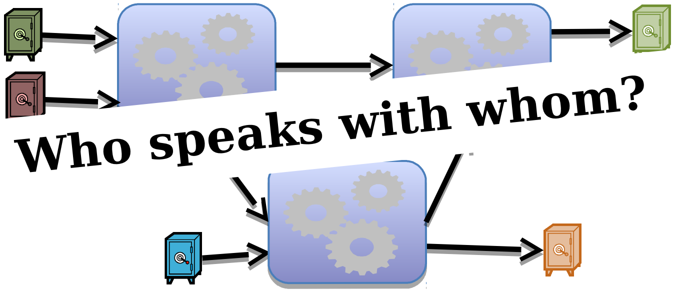
**Information will leak!!**

- And do not forget there is an adversary
  - not only observes public input/outputs of the system…
  - … also **knows** the privacy-preserving mechanism operation
  - e.g, ISP providers, system administrator, Data Retention, …

**How to quantify the information leaked?**

# This is a problem we all have
# Given an observation…

## Anonymous communications



Who speaks with whom?

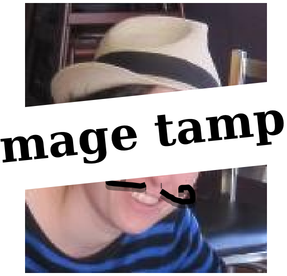
## Location privacy mechanisms

Which is the real location?

## Web traffic analysis countermeasures

Which web is this?

## Image forensics



Was the image tampered?

# Case study

# Anonymous communications

# Anonymous communications

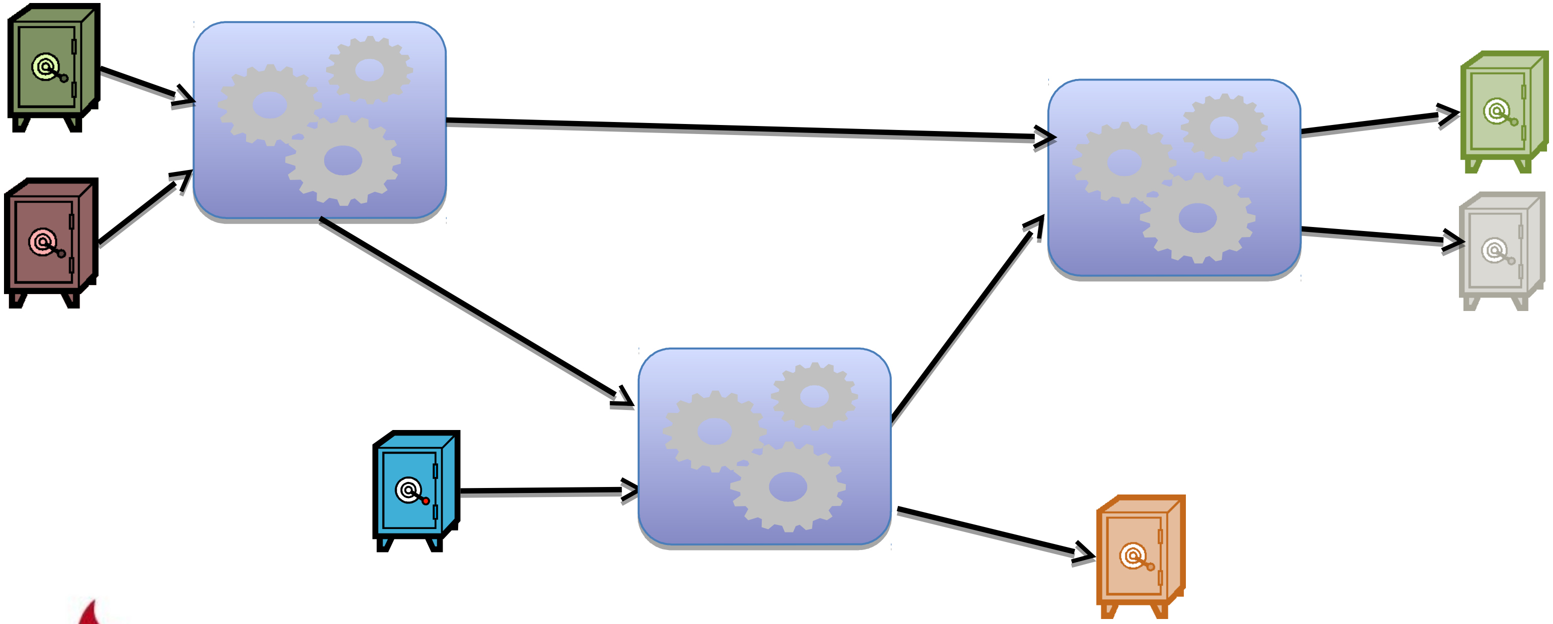
- Hide who speaks to whom
  - sender, receiver, type of service, network address, friendship network, frequency, relationship status.
- Main building block for privacy-preserving applications
  - Desirable privacy (comms, surveys,...)
  - Mandatory privacy (eVoting)
- Subject to constraints (bandwidth, delay,...)
  - They must leak information!

# Traffic analysis of Anonymous Communications

- Systems are evaluated against one attack at a time
  - Network constraints
  - Users knowledge
  - Persistent communications
  - …
- Based on heuristics and simplified models
  - Exact calculation of probability distributions in complex systems was considered as an intractable problem

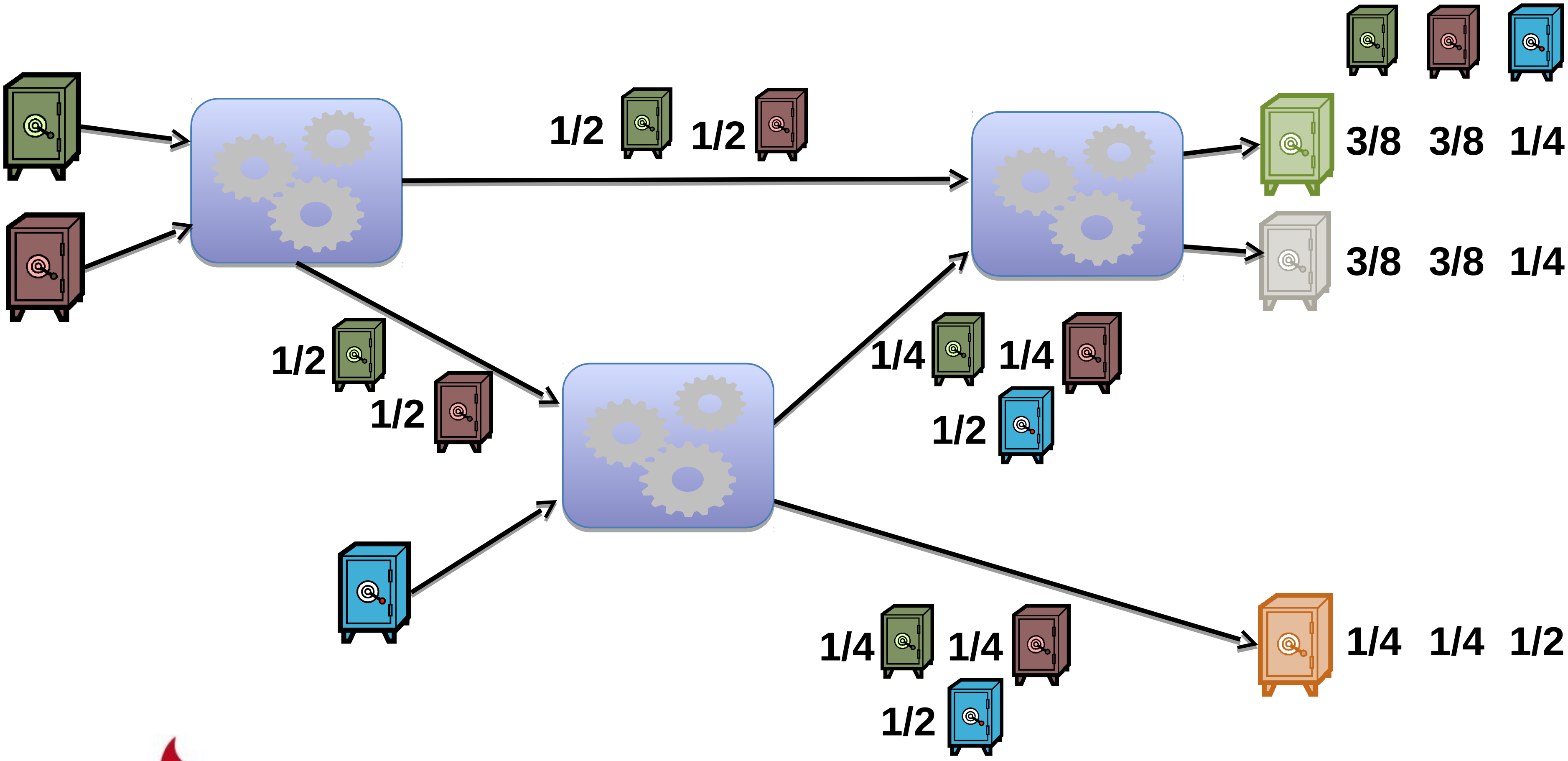
# Mix networks as an example

- Mixes hide relations between inputs and outputs
- Mixes are combined in networks in order to
  - Distribute trust (one good mix is enough)
  - Load balancing (no mix is big enough)



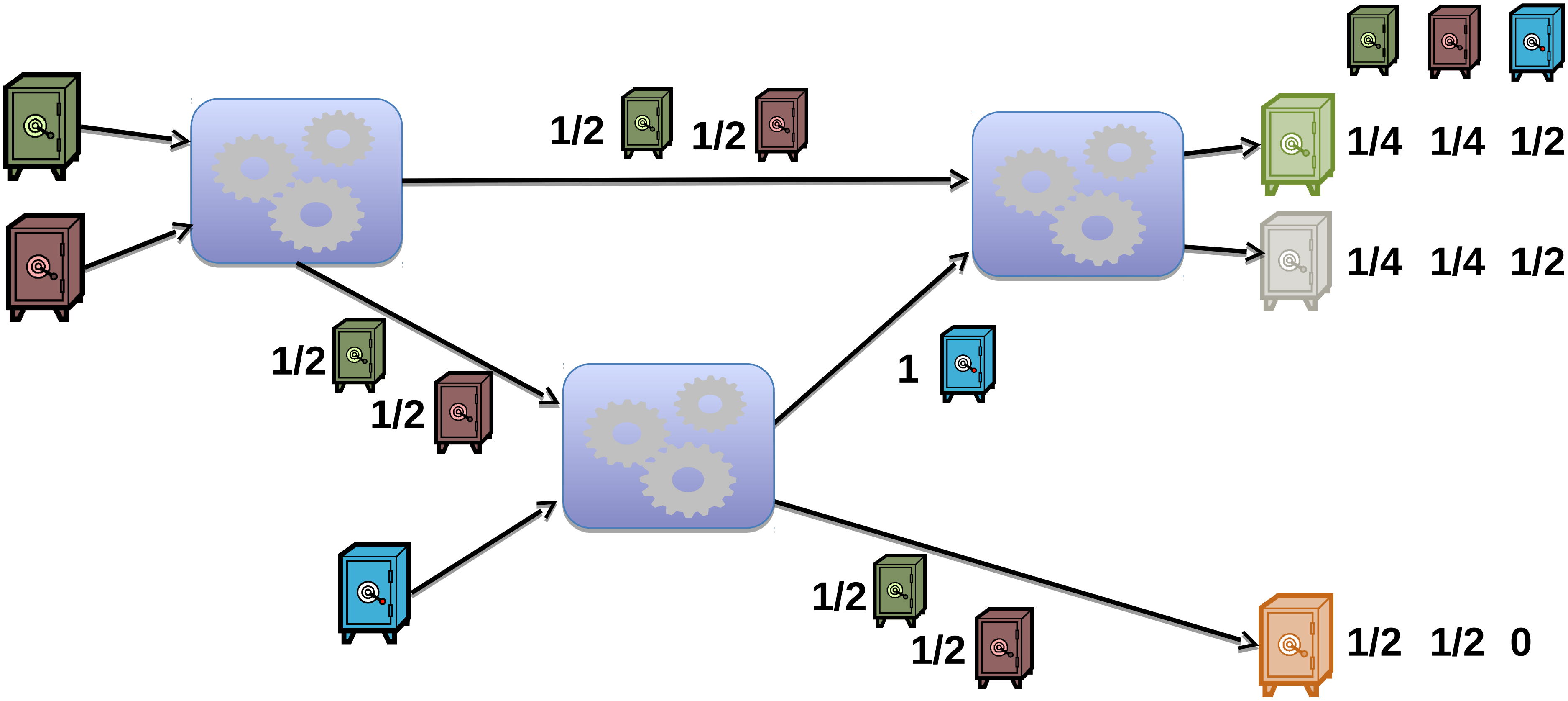
# The traffic analysis game

- Who speaks to whom?



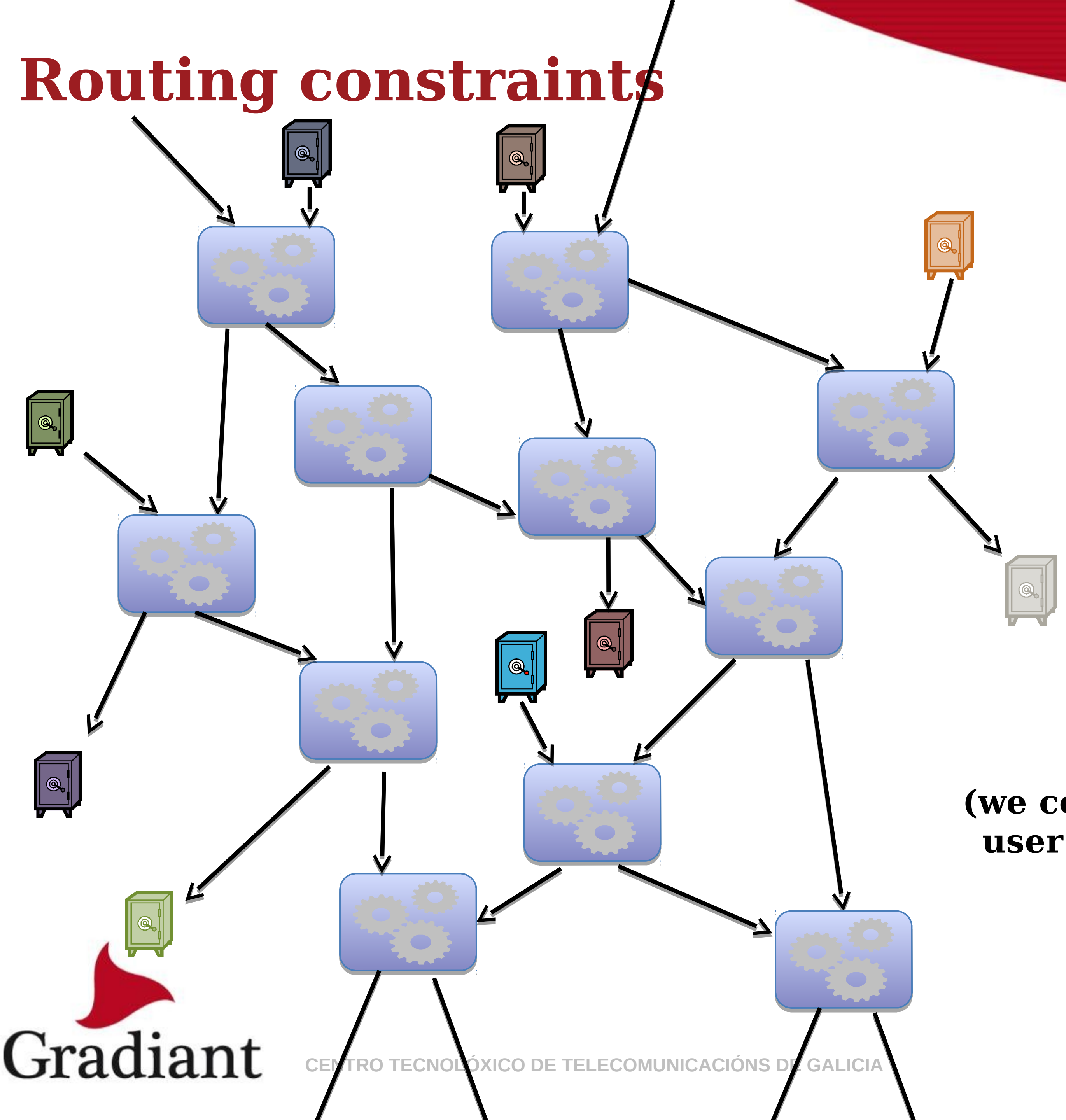
# Routing constraints

- Max Length = 2 hops



Non trivial given the observation!!

# Routing constraints

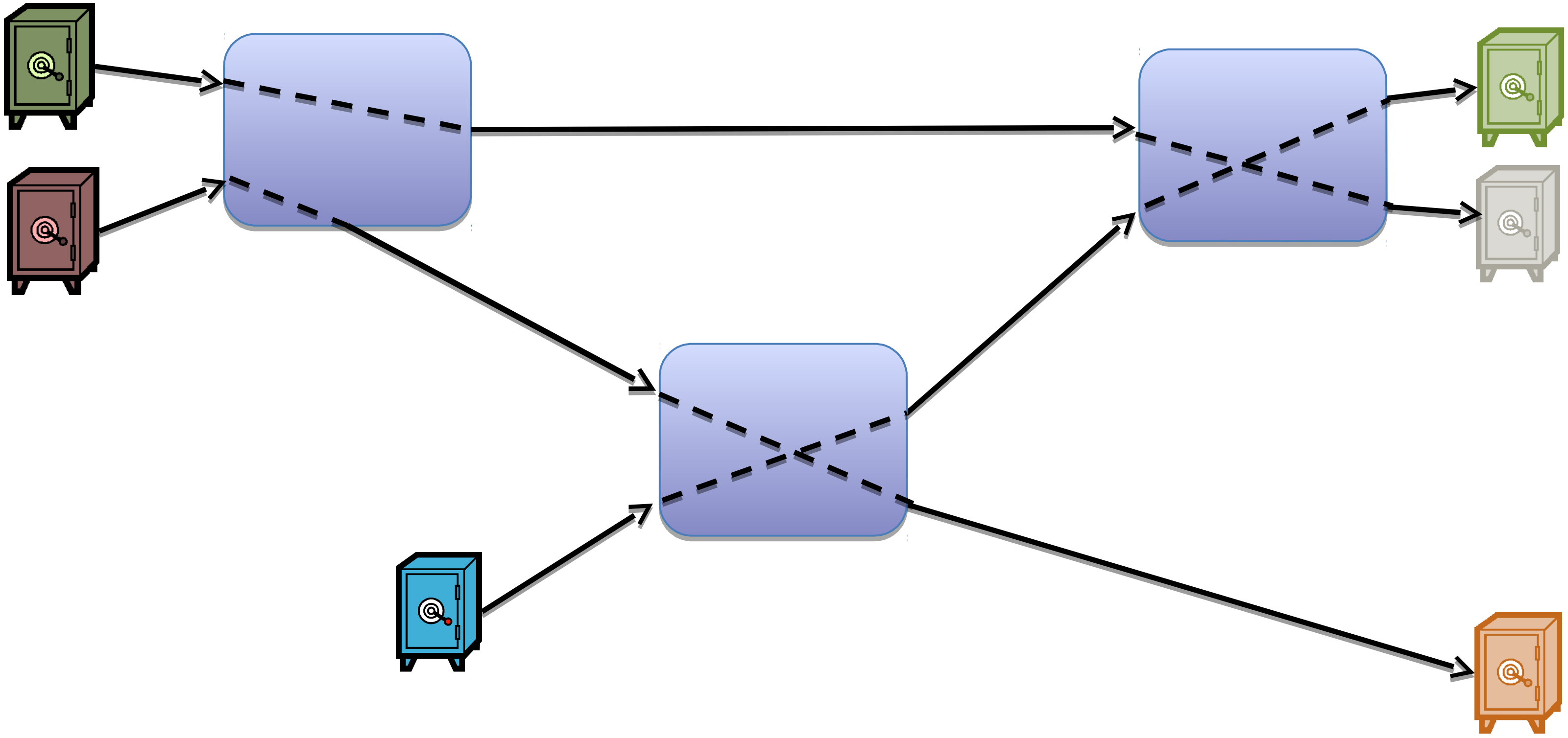


Really, non-trivial!

(we could think about user knowledge in the same way)

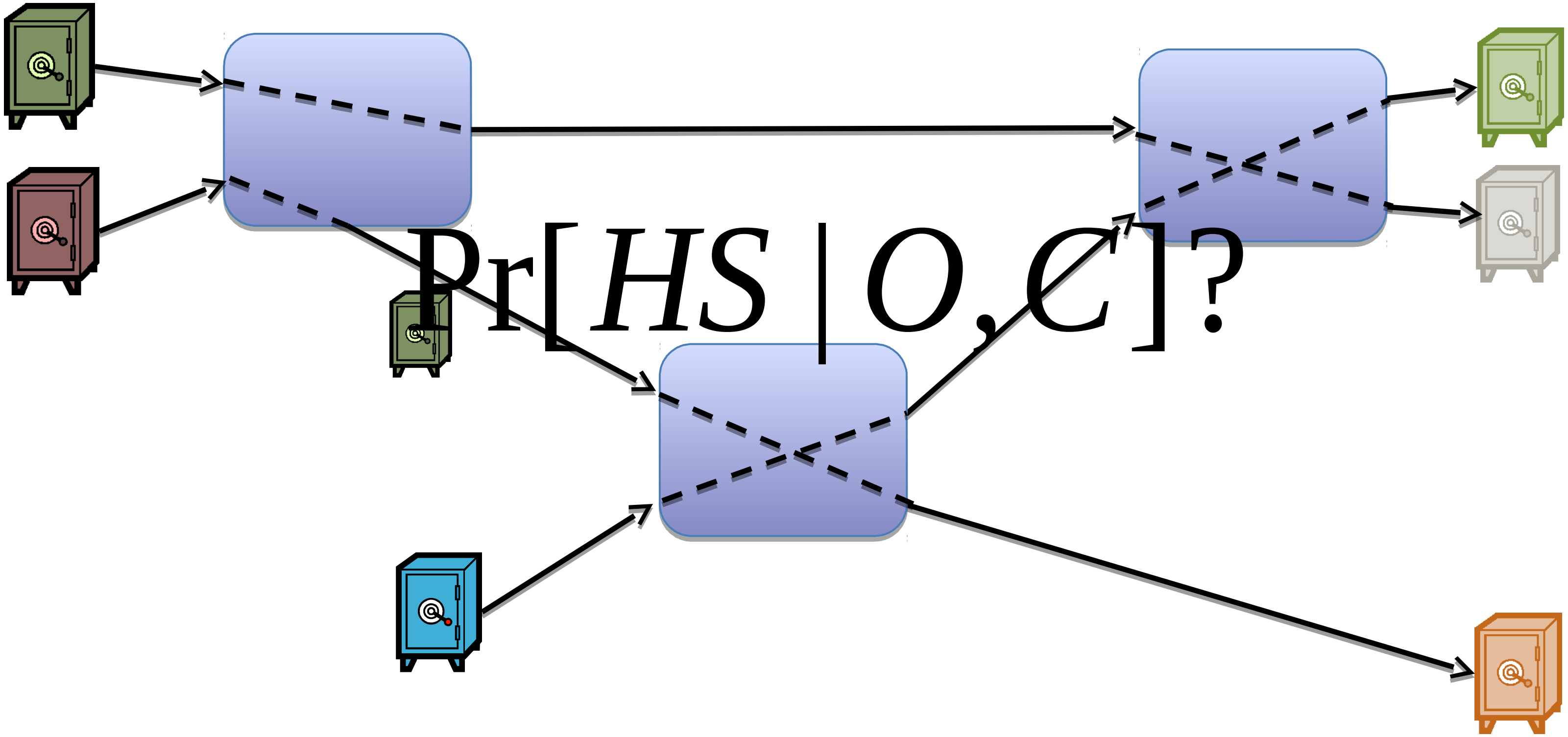
# (Re)Defining Traffic analysis

- Find hidden state of mixes



# (Re)Defining Traffic analysis

- Find hidden state of mixes

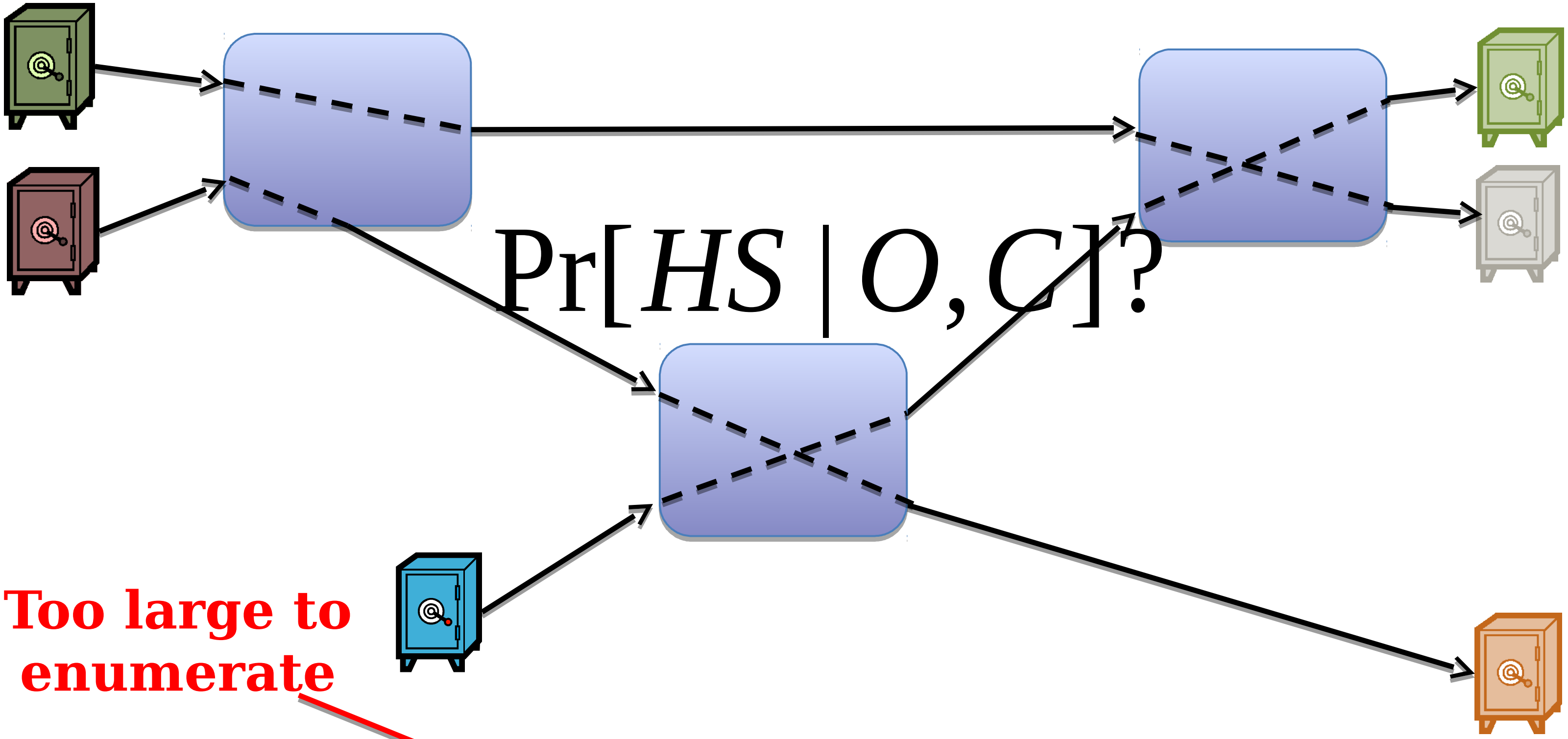


$$\Pr[HS \mid O,C]?$$

$$\Pr[HS \mid O,C] = \frac{\Pr[O \mid HS,C]\Pr[HS \mid C]}{\sum_{HS} \Pr[O \mid HS,C]}$$

# (Re)Defining Traffic analysis

- Find hidden state of mixes



$$\Pr[HS \mid O,C]?$$

**Too large to enumerate**

$$\Pr[HS \mid O,C] = \frac{\Pr[O \mid HS,C]\Pr[HS \mid C]}{\sum_{HS}\Pr[O \mid HS,C]} = \frac{\Pr[O \mid HS,C]K}{Z}$$

# Sampling to get probabilities

- Computing Pr[HS|O,C] infeasible: too many HS
  - … but we only care about marginal distributions
  - Is Alice speaking to Bob?
- if we had many samples of HS according to Pr[HS|O,C]
  - we could simply count how many times Alice speaks to Bob
- Markov Chain Monte Carlo methods
  - Sample from a distribution difficult to sample from directly

# Metropolis Hastings

- Simple
  1. Given $HS_0$ (an internal configuration of the mixes)
  2. Propose a new state $HS_1$
  3. Accept with probability min(1,α), reject otherwise

$$\alpha = \frac{\Pr[HS_1 \mid O, C] \cdot Q(HS_0 \mid HS_1)}{\Pr[HS_0 \mid O, C] \cdot Q(HS_1 \mid HS_0)} = \frac{\frac{\Pr[O \mid HS_1, C]\cancel{K}}{\cancel{Z}} \cdot Q(HS_0 \mid HS_1)}{\frac{\Pr[O \mid HS_0, C]\cancel{K}}{\cancel{Z}} \cdot Q(HS_1 \mid HS_0)}$$

- Pr[O|HS,C] is a generative model (in general simple)
- Q() is a proposal function
  - e.g., swap two links in a mix

**The stationary distribution corresponds to Pr[HS|O,S]**

**We can sample!**

The bayesian traffic analysis of mix networks,C. Troncoso and G. Danezis, 16th on Computer and Communications Security (CCS 2009)

# Why is this useful?

- Evaluation information theoretic metrics for anonymity

$$H = \sum_{R_i} \Pr[A \rightarrow R_i \mid O, C] \log(\Pr[A \rightarrow R_i \mid O, C])$$

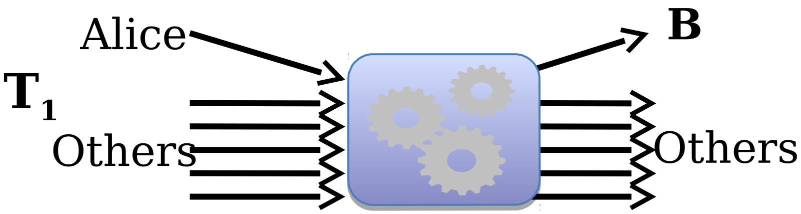
  - e.g., comparison of network topologies

- Estimating probability of arbitrary events
  - Input message to output message?
  - Alice speaking to Bob ever?
  - Two messages having the same sender?

- Accommodate new constraints
  - Key to evaluate new mix network proposals

Impact of Network Topology on Anonymity and Overhead in Low-Latency Anonymity Networks, C. Diaz, S. J. Murdoch, and C. Troncoso 10th Privacy Enhancing Technologies Symposium(PETS 2010)

# Persistent communications

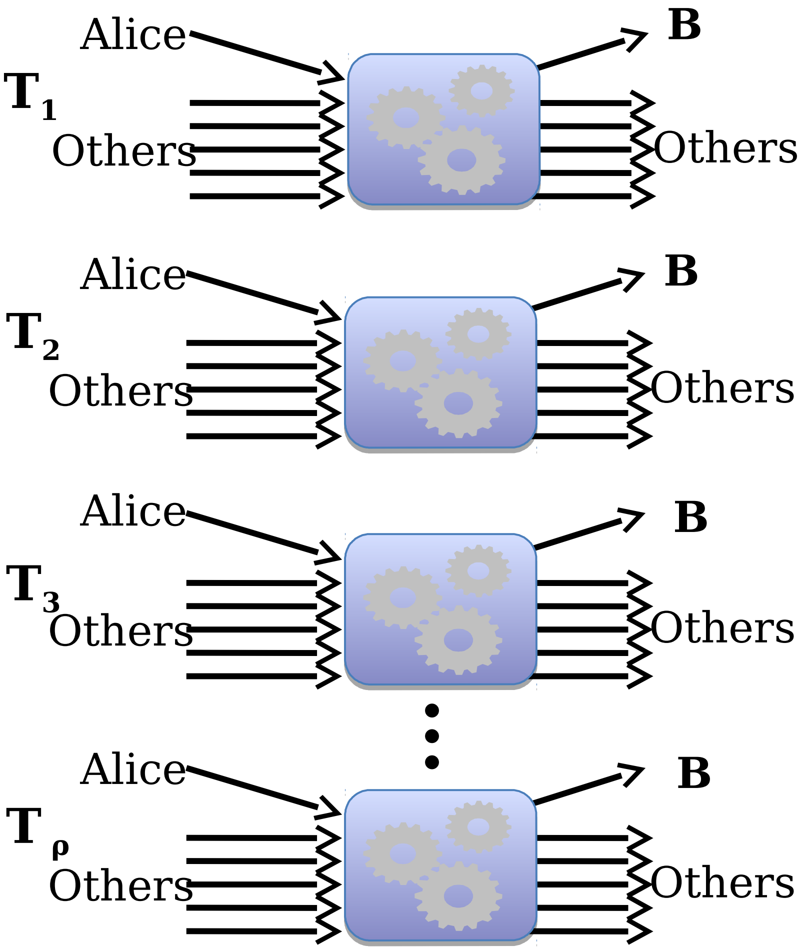


Perfect!

Anonymity set size = 6

Entropy metric $H_A$ = log 6

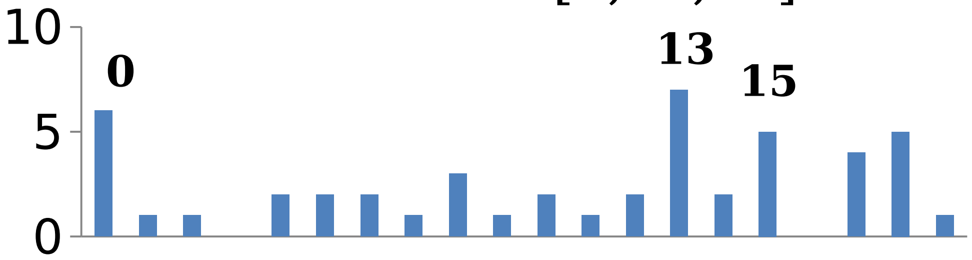
# Persistent communications



- Rounds in which Alice participates output a message to her friends
  - Her friends appear more often
  - We can infer set of friends!

# Statistical Disclosure Attacks

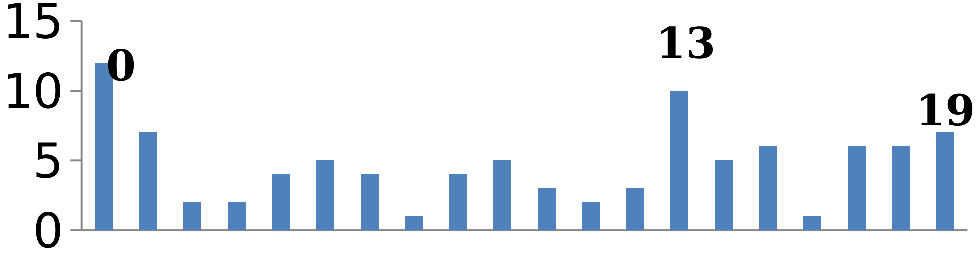
- Statistically find frequent receivers
  - Count & Substract “noise”
    - 20 users, 5 msgs/batch
    - Alice’s friends [0,13,19]



| Round | Receivers | SDA |
|---|---|---|
| 1 | [15, 13, 14, 5, 9] | [13, 14, 15] |
| 2 | [19, 10, 17, 13, 8] | [13, 17, 19] |
| 3 | [0, 7, 0, 13, 5] | [0, 5, 13] |
| 4 | [16, 18, 6, 13, 10] | [5, 10, 13] |
| 5 | [1, 17, 1, 13, 6] | [10, 13, 17] |
| 6 | [18, 15, 17, 13, 17] | [13, 17, 18] |
| 7 | [0, 13, 11, 8, 4] | [0, 13, 17] |
| 8 | [15, 18, 0, 8, 12] | [0, 13, 17] |
| 9 | [15, 18, 15, 19, 14] | [13, 15, 18] |
| 10 | [0, 12, 4, 2, 8] | [0, 13, 15] |
| 11 | [9, 13, 14, 19, 15] | [0, 13, 15] |
| 12 | [13, 6, 2, 16, 0] | [0, 13, 15] |
| 13 | [1, 0, 3, 5, 1] | [0, 13, 15] |
| 14 | [17, 10, 14, 11, 19] | [0, 13, 15] |
| 15 | [12, 14, 17, 13, | [0, 13, 17] |

# Statistical Disclosure Attacks

- Statistically finds frequent receivers
  - Count & Substract “noise”
    - 20 users, 5 msgs/batch
    - Alice’s friends [0,13,19]



- Efficient
- Needs a lot of data for reliability
- More complex models (replies, pool mixes, dummies)

| Round | Receivers | SDA |
|---|---|---|
| 1 | [15, 13, 14, 5, 9] | [13, 14, 15] |
| 2 | [19, 10, 17, 13, 8] | [13, 17, 19] |
| 3 | [0, 7, 0, 13, 5] | [0, 5, 13] |
| 4 | [16, 18, 6, 13, 10] | [5, 10, 13] |
| 5 | [1, 17, 1, 13, 6] | [10, 13, 17] |
| 6 | [18, 15, 17, 13, 17] | [13, 17, 18] |
| 7 | [0, 13, 11, 8, 4] | [0, 13, 17] |
| 8 | [15, 18, 0, 8, 12] | [0, 13, 17] |
| 9 | [15, 18, 15, 19, 14] | [13, 15, 18] |
| 10 | [0, 12, 4, 2, 8] | [0, 13, 15] |
| 11 | [9, 13, 19, 19, 15] | [0, 13, 15] |
| 12 | [13, 6, 2, 16, 0] | [0, 13, 15] |
| 13 | [1, 0, 3, 5, 1] | [0, 13, 15] |
| 14 | [17, 10, 14, 11, 19] | [0, 13, 15] |
| 15 | [12, 14, 17, 13, | [0, 13, 17] |

# Co-inferring routing and profiles

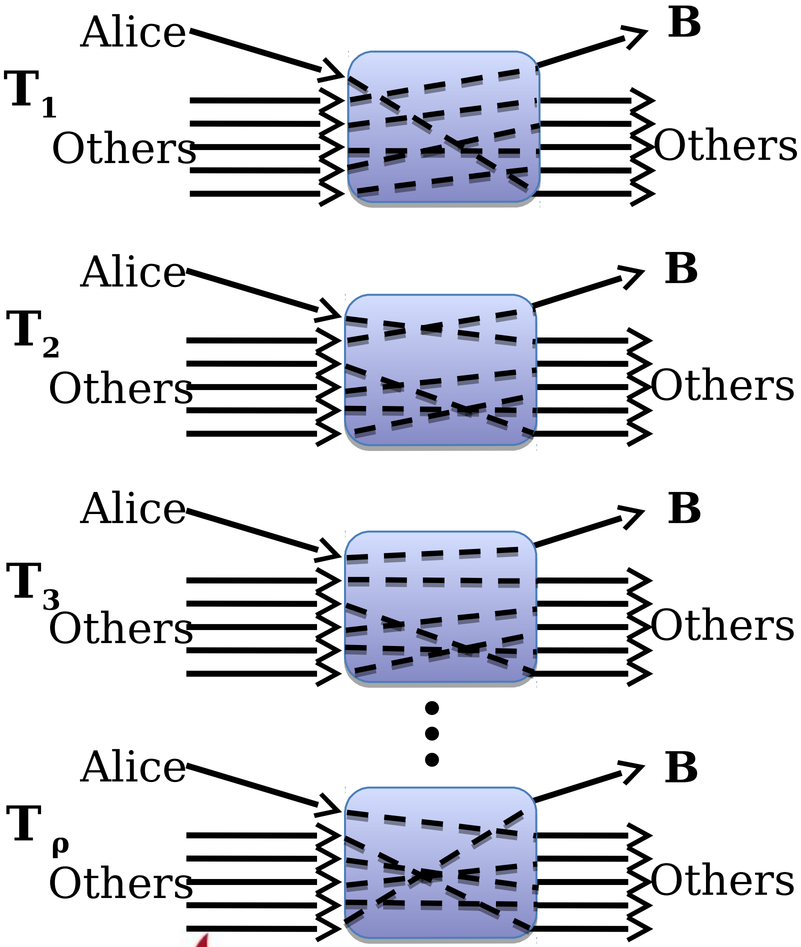
- A simple approach
  - Iterate profile and routing
  - Introduces systematic errors if done naively

- Actually we want to find $\Pr[M, \Psi \mid O, C]$
  - M is the routing, Ψ are the profiles (multinomial distribution)
  - Sounds familiar…

- Gibbs sampling
  - MCMC to sample from a joint distributions $\Pr[X, Y \mid O, C]$
  - Iterate $X \leftarrow \Pr[X \mid Y, O, C]$ and $Y \leftarrow \Pr[Y \mid X, O, C]$

Perfect matching disclosure attacks,C. Troncoso, B. Gierlichs, B. Preneel, and I. Verbauwhede. 8th International Symposium on Privacy Enhancing Technologies (PETS 2008)

# Gibbs sampling for anonymity systems



## From matching to profiles

$$\Pr[\Psi \mid M, O, C]$$

**Observation**

$V_{AB} = 1$ $V_{AO} = 3$

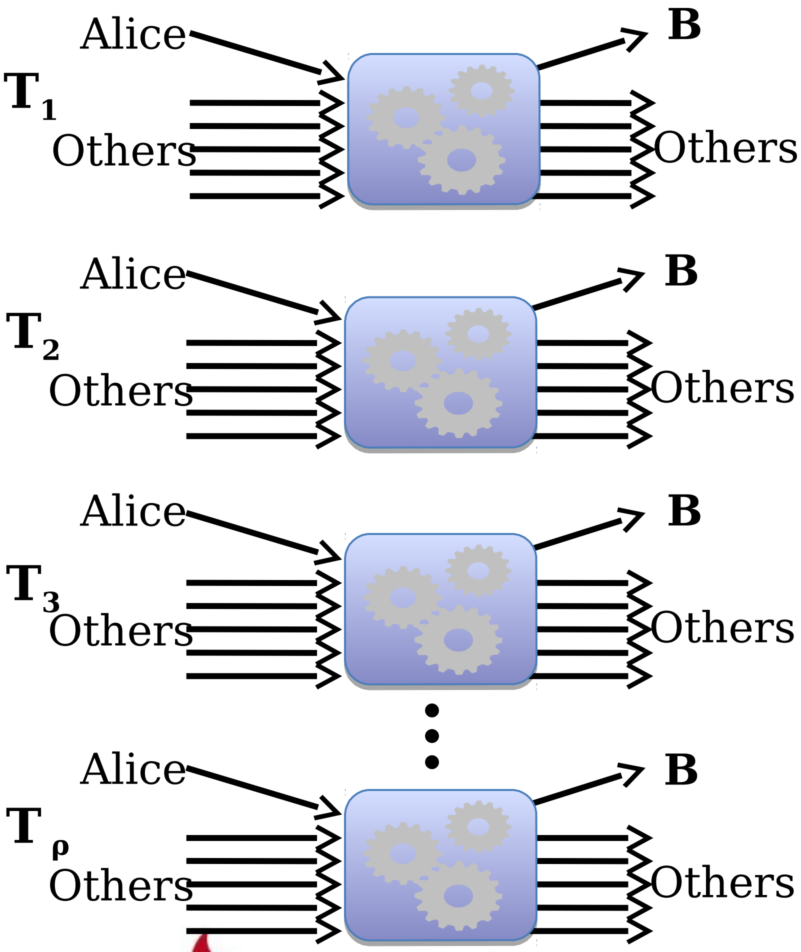
$V_{OB} = 3$ $V_{OO} = 17$

**Count messages and use the multinomial prior**

$$\Psi = \text{Dirichlet}\,(V_{AB}, V_{AO})$$

CEN Vida: How to use Bayesian inference to de-anonymize persistent communications. George Danezis, and Carmela Troncoso, 9th Privacy Enhancing Technologies Symposium (PETS 2009).

# Gibbs sampling for anonymity systems



## From profiles to matchings

$$\Pr[M \mid \Psi, O, C]$$

$$\Psi_{Alice} = \{\Pr[A \rightarrow B], \Pr[A \rightarrow O]\}$$

$$\Psi_{Others} = \{\Pr[O \rightarrow B], \Pr[O \rightarrow O]\}$$

**Sadly not as simple…**

1. If possible analytical
2. Use MCMC-MH
3. Other alternatives?

# And if profiles are dynamic?

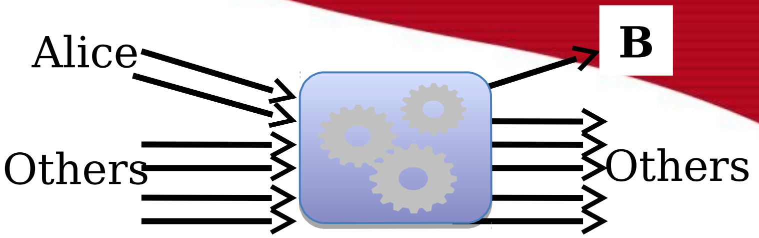
- Previous methods work for static behavior
  - But this does not seem very realistic…

- The Bayesian approach: Particle filtering
  - Sequential Monte Carlo
  - Infer dynamic hidden variables when the state space is intractable analytically

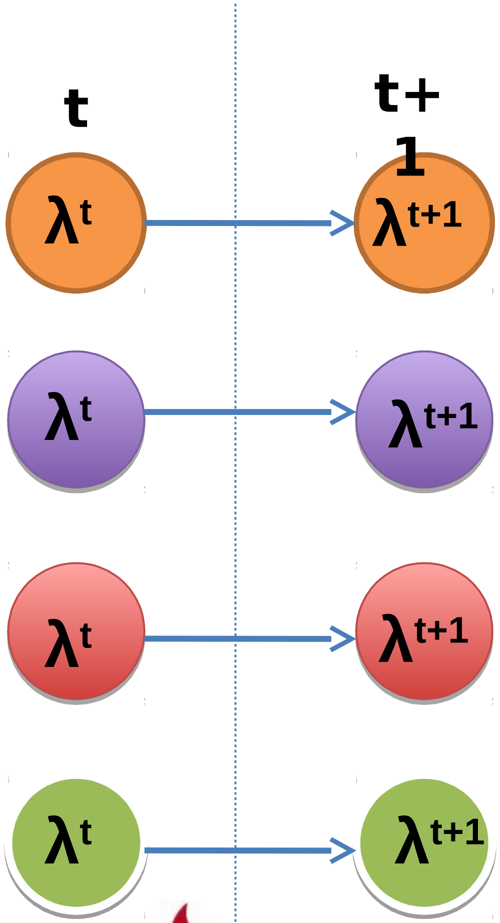
- The adversary observes volumes of communication and wants to infer poisson rates that generates them

$$\Pr[\lambda_{AB_t} \mid \lambda_{AB_{t-1}}, O, C]$$

# Toy example



$V_A=2,\ V_O=4,\ V_B=1,\ V_{OO}=5$

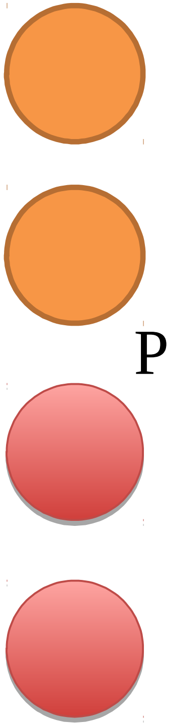


**Weight particles:**

i. **Likelihood**

ii. **Evolution**

iii. **Proposal**



$$\Pr[(\lambda^t_{AB}, \lambda^t_{OB}) \,|\, V_*]$$

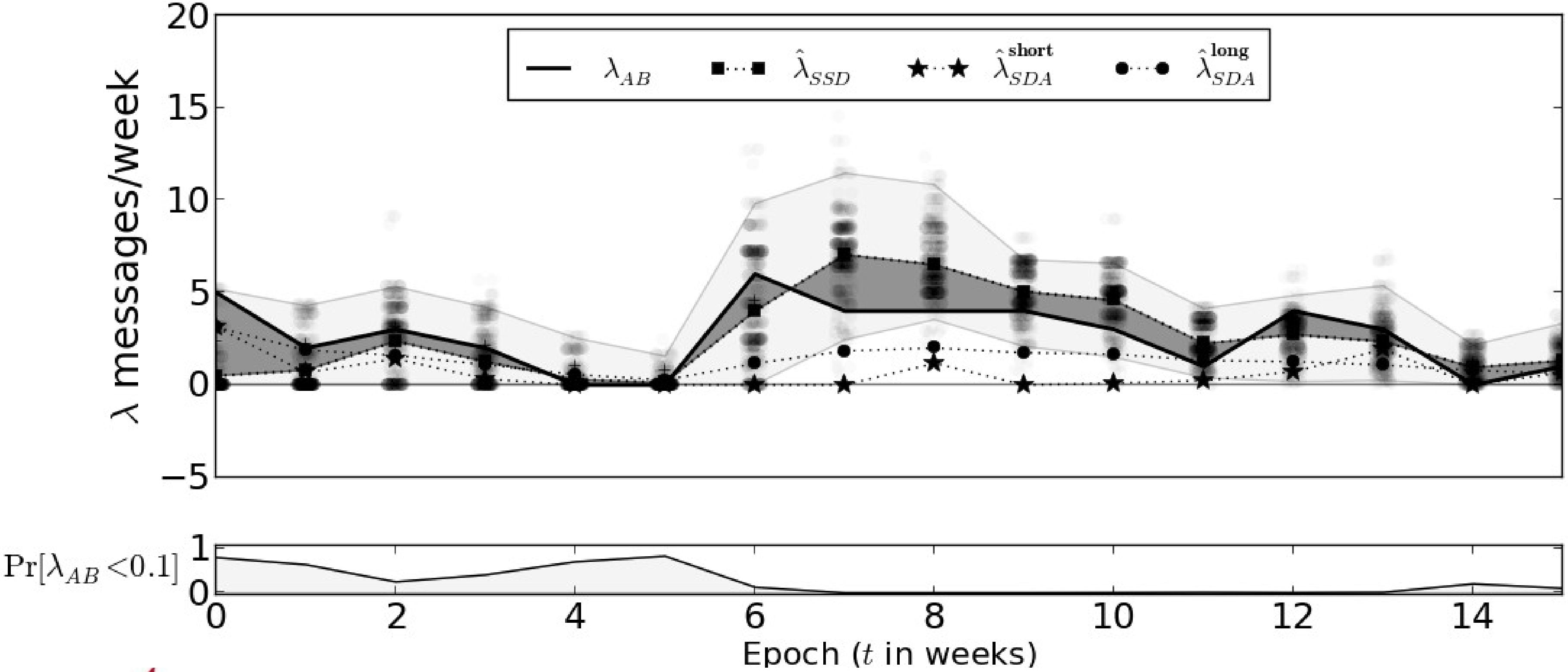
**1. Propose new particles**

**2. Likelihood given Obs and previous state**

**3. Re-sample**

# Results

- Enron dataset (http://www.cs.cmu.edu/~enron/)



# Advantages

- Systematic
  - Generative model tends to be easy

- Return probability distributions
  - More informative than Maximum Likelihood
  - Allow for multiple inferences

- Confidence estimates
  - Key in real analysis!

- **What I did not say**
  - **I have avoided all the scary details**
  - **Getting the model correctly is non-trivial**

# Applications

- We have seen three Bayesian methods
  - Metropolis Hastings sampling Pr[HS|O,C]
    - Location privacy - tracking
    - Differential privacy
  - Gibbs sampling Pr[X,Y|O,C]
    - Location privacy – de-anonymization
  - Particle filtering $\Pr[\lambda_t|\lambda_{t+1},O,C]$
    - Privacy-preserving video surveillance

- Lots to do
  - Tor: website fingerprinting, flow correlation, flow watermarking, routing,…
  - Location privacy: dynamic behaviour
  - Cloud computing: side channels

# The message I wanted to convey

- We are solving the same problem again and again
- Bayesian inference as systematic approach
  - Allows to tackle complex scenarios
  - Sampling reduces computational requirements

# Thanks!

I hope I have awaken your curiosity ⏪

**ctroncoso@gradiant.org**