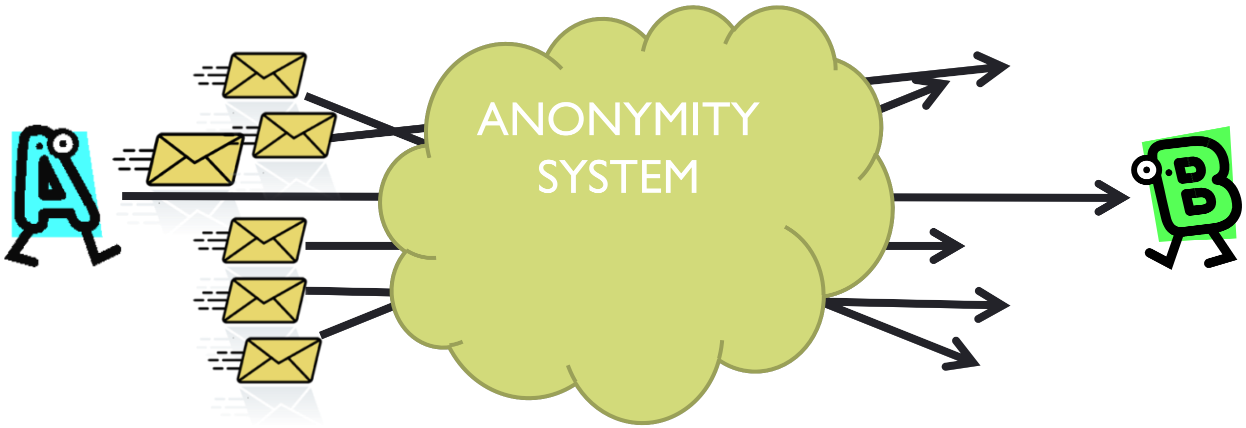
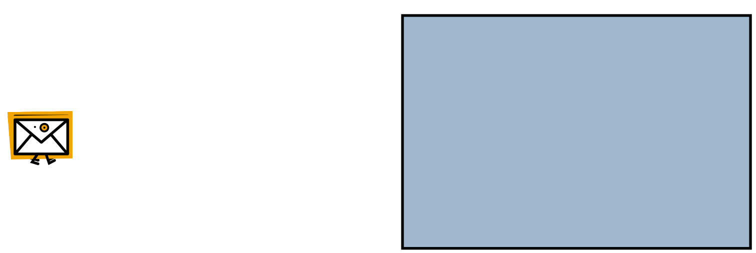
# **Vida**:
# **B**ayesian **I**nference to **D**e-**A**nonymize communications

**Carmela Troncoso**, KU Leuven/COSIC
(Intern at Microsoft Research Cambridge Sept-Nov 2008)
George Danezis, Microsoft Research Cambridge

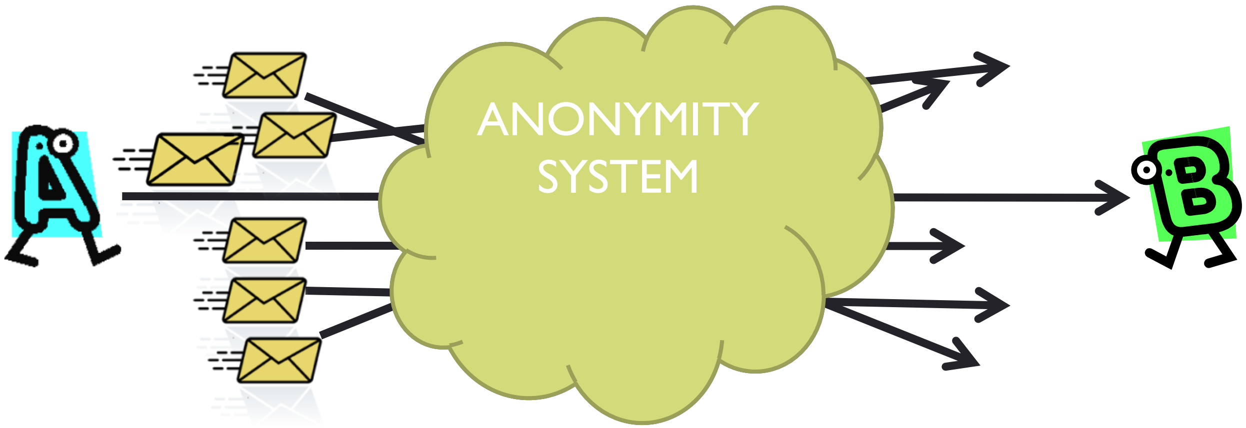
# Anonymous communications



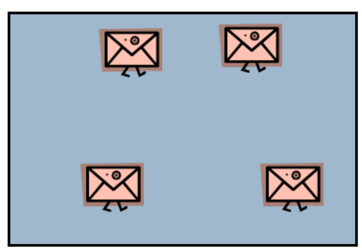

- Privacy, e-voting, protection of trade secrets, high security military appplications
- The Threshold Mix [Chaum81]



# Anonymous communications




- Privacy, e-voting, protection of trade secrets, high security military appplications
- The Threshold Mix [Chaum81]



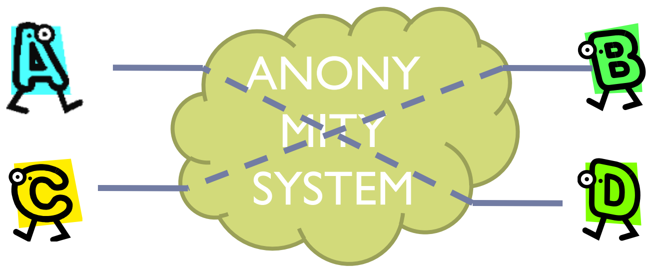
# Traffic analysis: Intersection attacks

- Exploit persistent patterns for de-anonymization
  - Disclosure Attack [Kesdogan et al 02]
    - Set theory
    - NP-Problem
  - Statistical Disclosure Attack [Danezis 03]
    - Computationally feasible
    - Inaccurate
  - Perfect Matching Disclosure Attack [Troncoso et al 08]
    - Perfect matching
    - Reuse for profiles

- Ad-hoc studies,  difficult to estimate errors
  - Bayesian inference to de-anonymize and profile systematically

# Redefining the traffic analysis problem

- Find *"hidden state"* of an anonymity system




$$\Pr(HS \mid O, C)$$

- If we apply Bayes theorem…

$$\Pr(HS \mid O, C) = \frac{\Pr(O \mid HS, C)\,\Pr(HS \mid C)}{\sum_{HS} \Pr(HS, O \mid C)}$$

Prior information

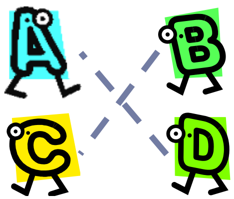
Too large to enumerate!!

$$HS_1, HS_2, HS_3, \ldots \sim \Pr(HS \mid O, C)$$

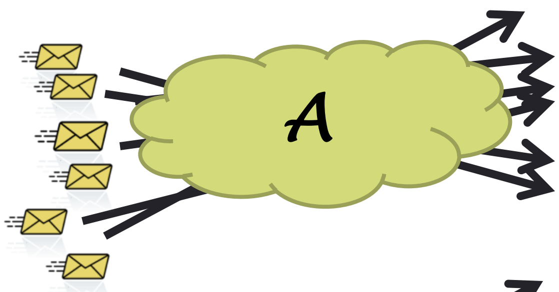
- Markov Chain Monte Carlo methods

# The Vida Black-box model

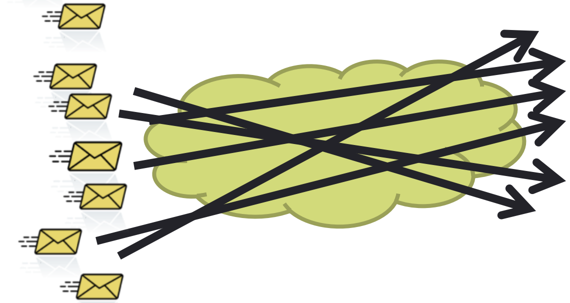
**$\Psi$ – User Profiles**



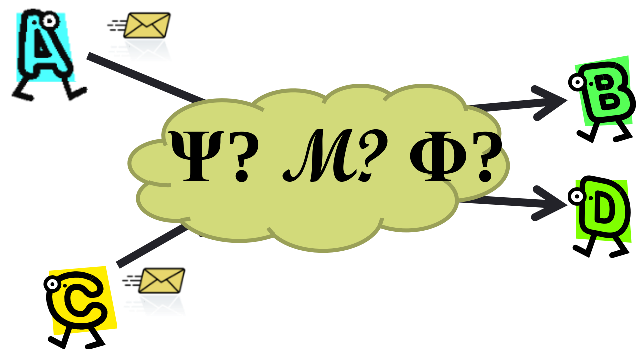
**$\mathcal{M}$ – Sender-Receiver match**




**$\mathcal{A}$ – Anonymity system**



**$\Phi$ – Assignment**




**$\mathcal{O}$ – Observation**

$$\Pr(M,\Phi,\Psi \mid O,A) = \frac{\Pr(M \mid \Psi)\cdot\Pr(\Phi \mid A)}{\Pr(O \mid A) \equiv Z}\Pr(\Psi \mid A)$$

Prior information

**(full derivation in the paper)**

# Markov Chain Monte Carlo Methods

- Sample from a distribution difficult to sample from directly

$$\Pr(M,\Phi,\Psi \mid O,A) = \frac{\Pr(M \mid \Psi)\cdot\Pr(\Phi \mid A)}{\Pr(O \mid A) \equiv Z}$$

- Constructs a Markov Chain with stationary distribution equal to the target distribution

- Gibbs sampling
  - Efficient for sampling joint distributions
  - Eliminate the need to compute Z

# Gibbs sampling for Vida

$$\Pr(M, \Phi, \Psi \mid O, A)$$

- Iterative ly draw samples from the marginal distributions

$$\Phi_j, M_j \sim \Pr(\Phi, M \mid \Psi_{j-1}, O, A)$$

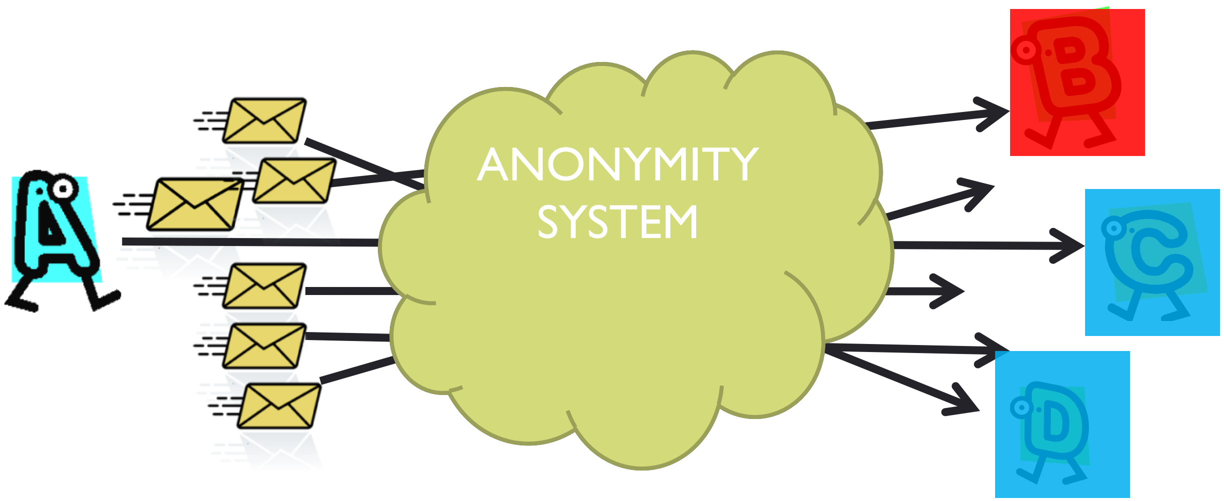
$$\Psi_j \sim \Pr(\Psi \mid \Phi_j, M_j, O, A)$$

  - **Φ, $\mathcal{M}$** – Find perfect matching (reject if not valid)
  - **Ψ**– Use the dirichlet distribution (prior of multinomial)

$$\Psi_A \sim Dirichlet(Ct_M(A \rightarrow B)+1, Ct_M(A \rightarrow C)+1, \ldots, Ct_M(A \rightarrow Z)+1)$$

# Simple Vida: Red-Blue Model

- Do we actually want to know to whom every user speaks?
  - Who sent a message to Bob?
  - Who is friends with receiver Bob?




  - Profiles become binomial (Red or Blue)
  - Blue receivers are equivalent when making assignments $\mathcal{M}$

# Evaluation

- Synthetic anonymized traces

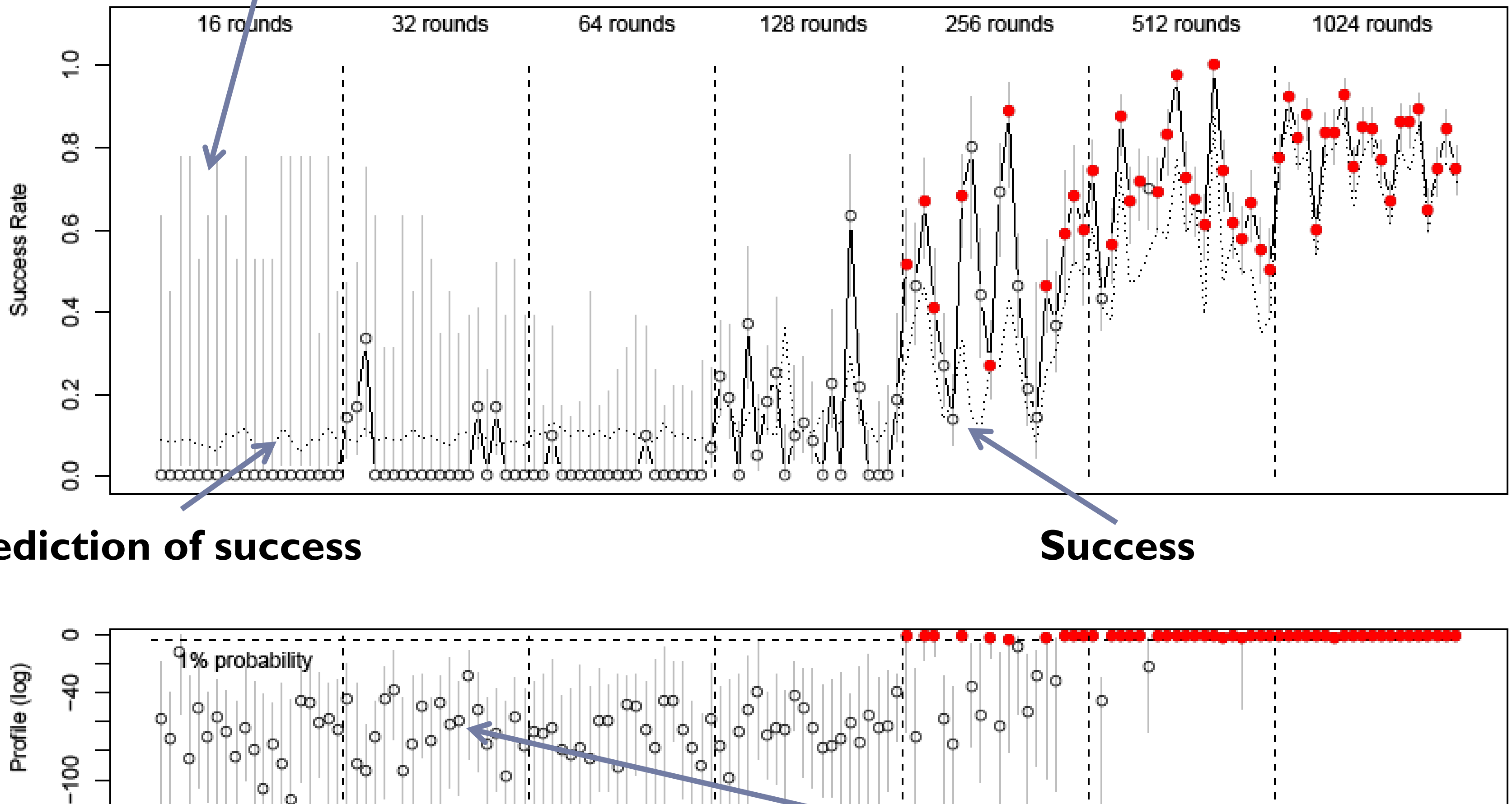
| | |
|---|---|
| **Users** | **1000** |
| **Friends** | **5** |
| **Threshold** | **100** |

- Target sender in 20% of the rounds
  - Friend of Red receiver
  - Allow profiling of other users

- Use Gibbs sampler to guess receiver (200 samples)
  - Prior belief $\Pr(\Psi \mid A)$ Beta(0.01,0.01)
  - Bayes optimal criterion

# Success rate

**90% confidence interval**

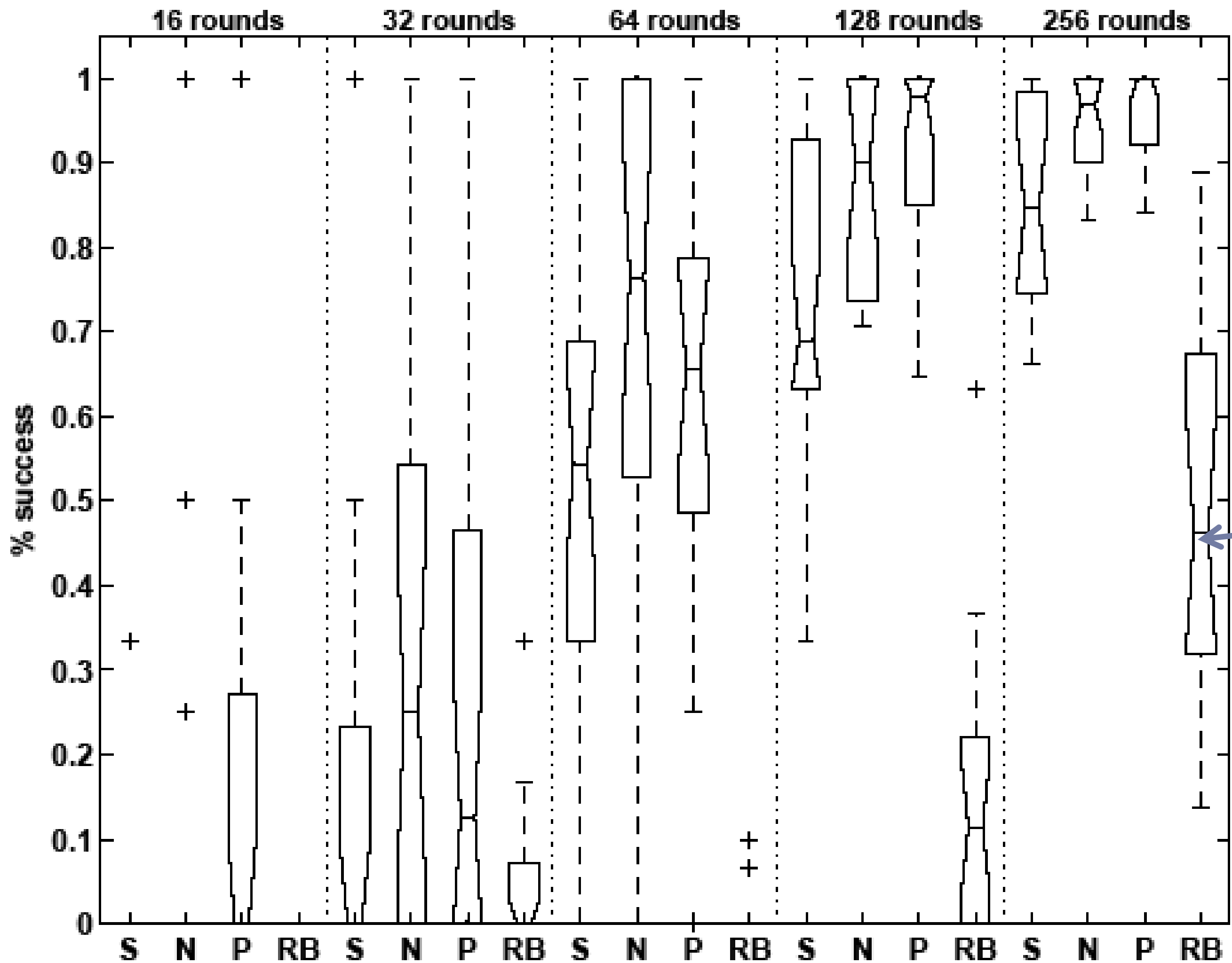



**Prediction of success**

**Success**

**Profile quality**

# Comparison with previous work




- S-SDA
- N-NSDA
- P-PMDA
- RB-Vida

**Bayes optimal criterion does not guarantee a perfect matching!**

**Good error estimates**

**Full distribution**

# Lots to do…

- Weighted incomplete Bipartite graph
  - Threshold mix is easy to compute
  - What about networks?
    - **The Bayesian Traffic Analysis of Mix Networks.** Carmela Troncoso and George Danezis. CCS 09
- Increase constraints on the profiling
  - Modeling more difficult but better results
- Social Networks inference
  - Prior information can be easily added to the model
- Beyond communications
  - Location privacy, Database de-anonymization

# Conclusions

- Vida Black-Box model
  - Generic
  - Accommodates any anonymity system
  - No need to know number of friends

- Vida Red-Blue model
  - Efficiently de-anonymizes targeted senders/receivers

- Markov Chain Monte Carlo as basis for traffic analysis
  - 3 Key advantages:
    - Requires generative model
    - Good estimation of errors
    - Systematic

# Questions?

Thanks for your attention!!

Carmela.Troncoso@esat.kuleuven.be