Universidad de Vigo
GPSC
Signal Processing in Communications Group

Meet the family of Statistical Disclosure Attacks

Simon Oya, Carmela Troncoso and Fernando Pérez-González

Gradiant
CENTRO TECNOLÓXICO DE TELECOMUNICACIÓNS DE GALICIA

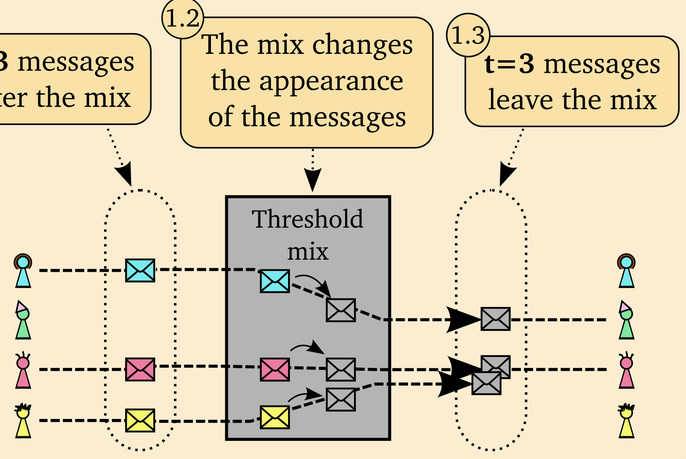
System Model and Problem Statement

1st **N=4** users communicate using an annonymous channel denominated threshold mix.

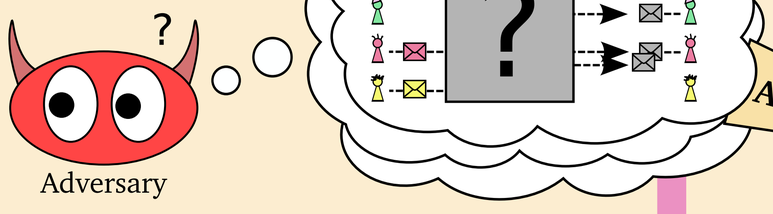
1.1 **t=3** messages enter the mix

1.2 The mix changes the appearance of the messages

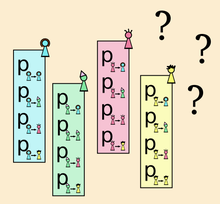
1.3 **t=3** messages leave the mix



2nd An adversary observes the incoming and outgoing messages several times

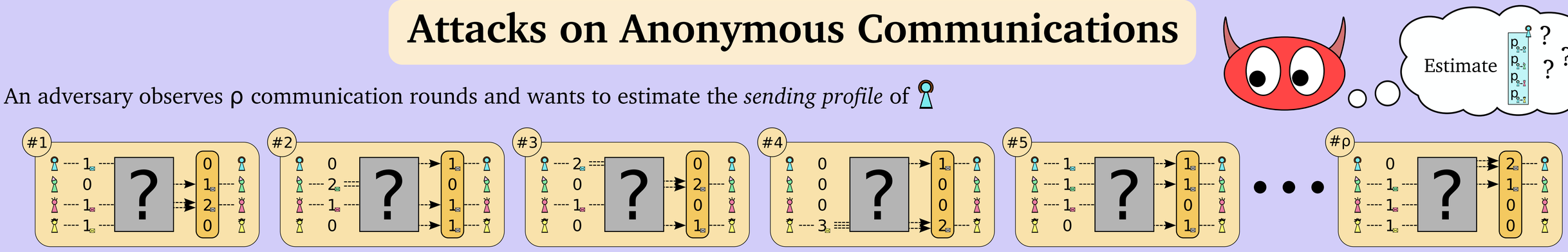


3rd The adversary wants to infer the *sending profiles* of the users



Attacks on Anonymous Communications

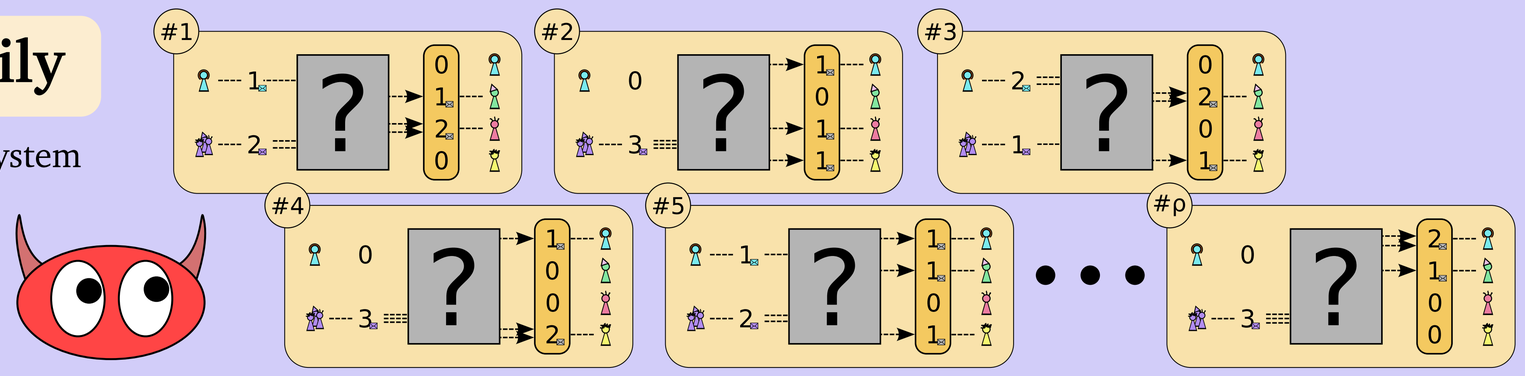
An adversary observes ρ communication rounds and wants to estimate the *sending profile* of



Statistical Disclosure Attack family

SDA's idea: separate from the rest of the users:

Equivalent system for the SDA adversary:

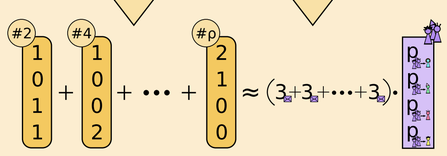


SDA0 (2003)

1st Estimate the *sending profile* of all users but together:

Rounds where does not participate

Messages sent by in those rounds

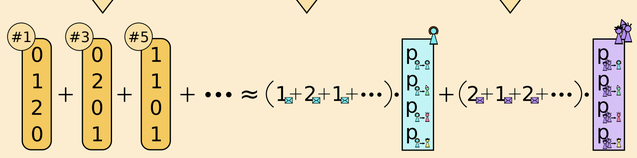


2nd Estimate 's *sending profile*

Rounds where participates

Messages sent by in those rounds

Messages sent by in those rounds



SDA1 (New!)

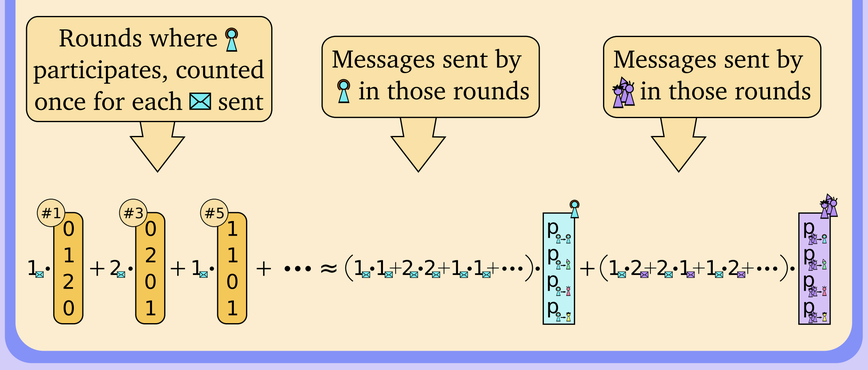
1st Estimate the *sending profile* of as in SDA0

2nd Estimate 's *sending profile*, now exploiting the information about in the rounds

Rounds where participates, counted once for each sent

Messages sent by in those rounds

Messages sent by in those rounds



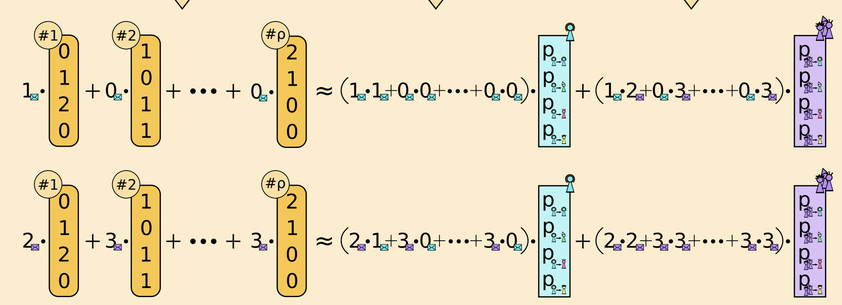
SDA2 (New!)

Estimate both 's and 's *sending profiles* simultaneously, using information from **all rounds**, following the approach of SDA1

All rounds, counted once for each ( ) sent

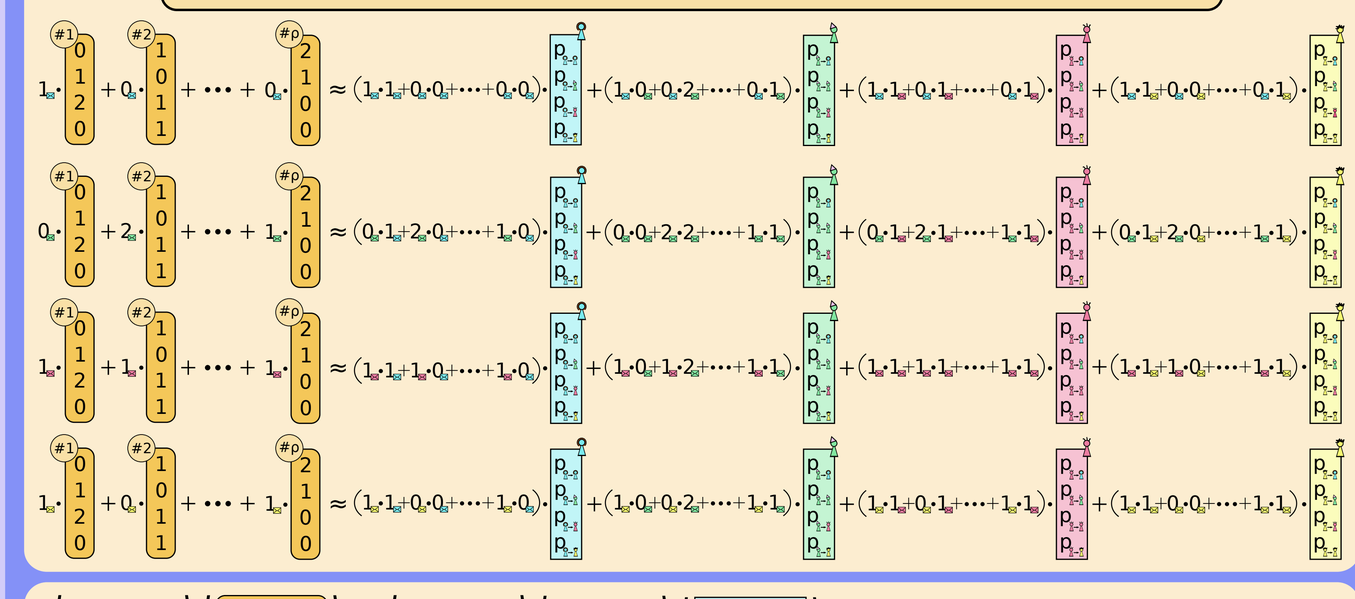
Messages sent by in those rounds

Messages sent by in those rounds



Least Squares Disclosure Attack (LSDA) (2012)

Estimate the *sending profiles* of **all users** simultaneously, using information from **all rounds**



$$\mathbf{Y} \approx \mathbf{U}\mathbf{P} \quad \Rightarrow \quad \hat{\mathbf{P}}^{\mathtt{LSDA}} = \left(\mathbf{U}^T\mathbf{U}\right)^{-1}\mathbf{U}^T\mathbf{Y}$$

Attack Performance

Mean Squared Error in the estimation of the sender profile of :

$$\mathrm{MSE}^{\mathtt{LSDA}} \approx \frac{1}{\rho}\left\{\left(f_{\text{}}^{-1}-1\right)\left(1-\frac{1}{t}\right)\bar{\mu}_{\mathtt{LSDA}}+\frac{f_{\text{}}^{-1}}{t}\cdot\mu_{\text{}}\right\}$$

$$\mathrm{MSE}^{\mathtt{SDA2}} \approx \frac{1}{\rho}\left\{\left(f_{\text{}}^{-1}-1\right)\left(1-\frac{1}{t}\right)\bar{\mu}_{\mathtt{SDA2}}+\frac{f_{\text{}}^{-1}}{t}\cdot\mu_{\text{}}\right\}$$

Rounds observed

's *sending frequency* (from 0 to 1)

Mix *threshold*

$\mu_{\text{}} = 1 - \sum_{i=1}^{} p_{\text{}\to i}^2$

's *uniformity*: from 0 to ≈1

$\bar{\mu}_{\mathtt{LSDA}} \leq \bar{\mu}_{\mathtt{SDA2}}$

Experimental results:

ρ = 20.000 rounds
N = 100 users
f_i = 1/N
t = 10 messages