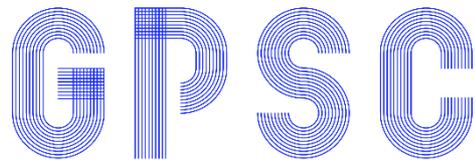


# Understanding the Effects of Real-World Behavior in Statistical Disclosure Attacks

---

**Simon Oya**, Carmela Troncoso and Fernando Pérez-González

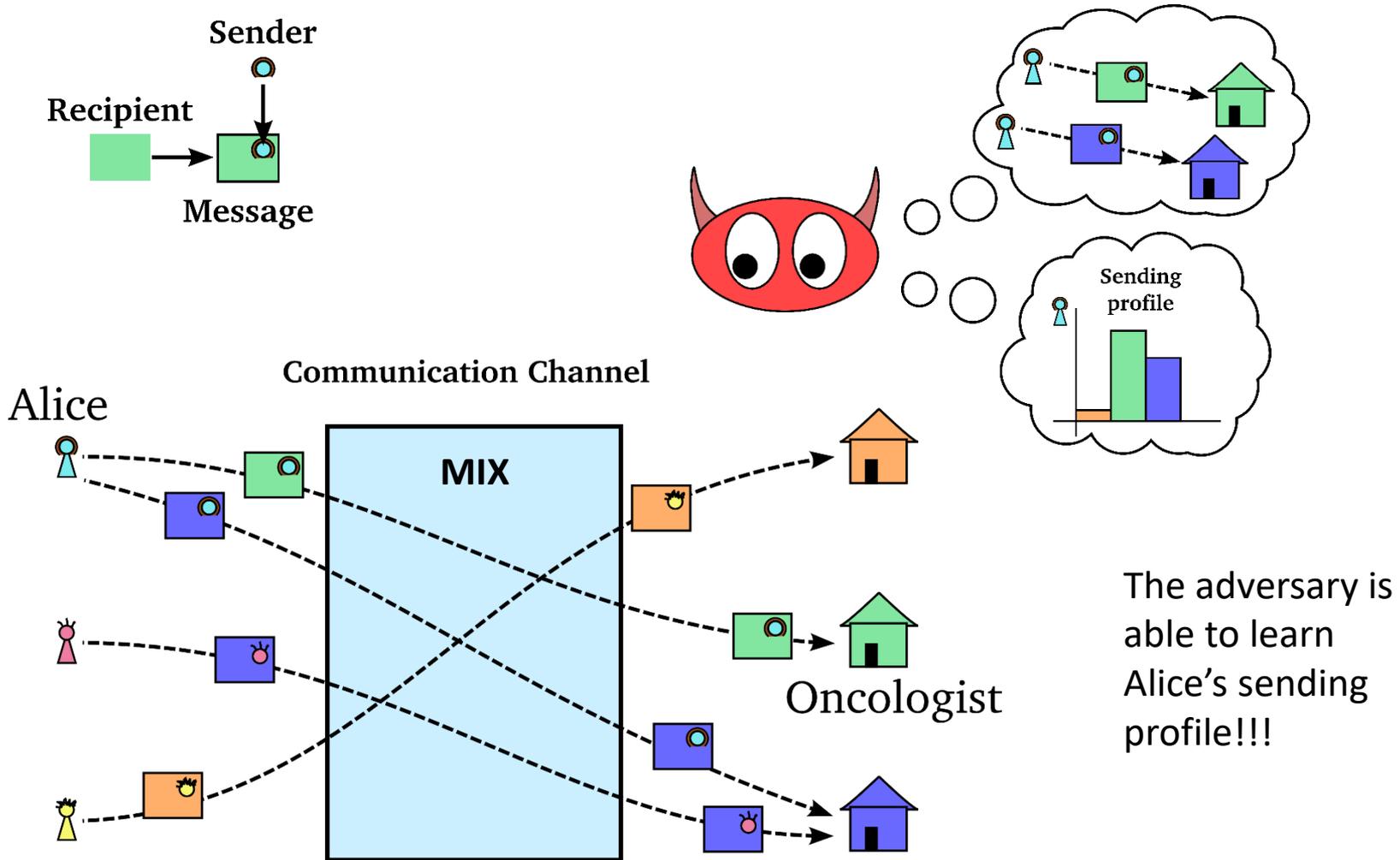
Universidade de Vigo



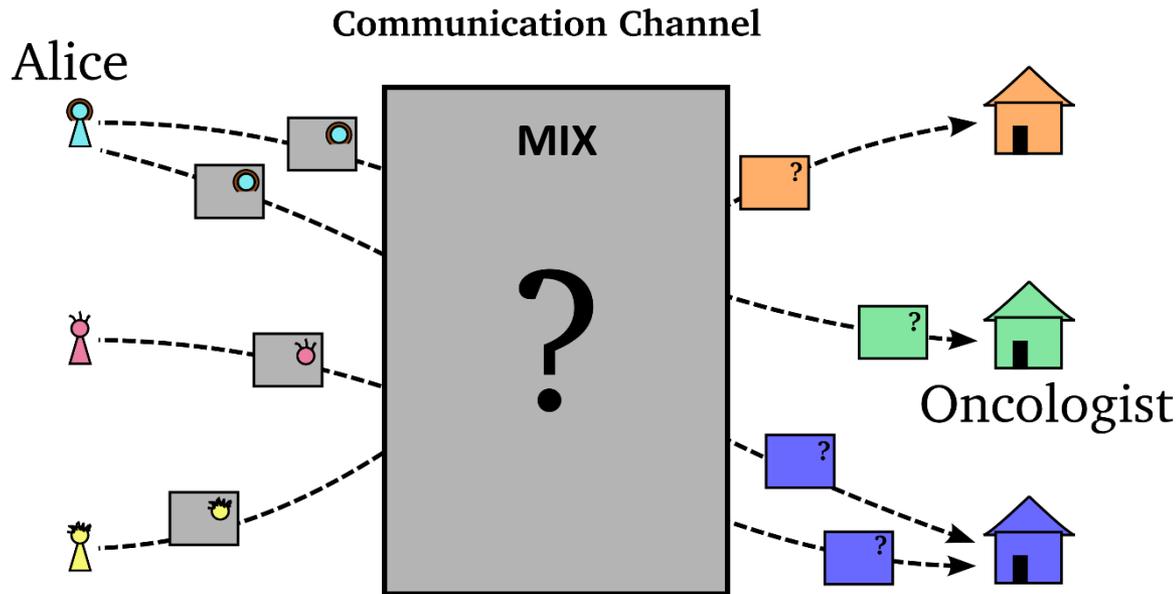
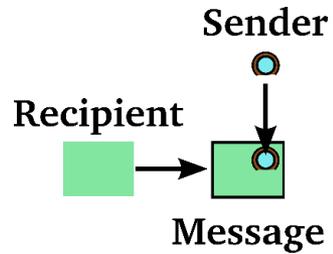
Signal Processing in  
Communications Group



# Introduction. Mixes.



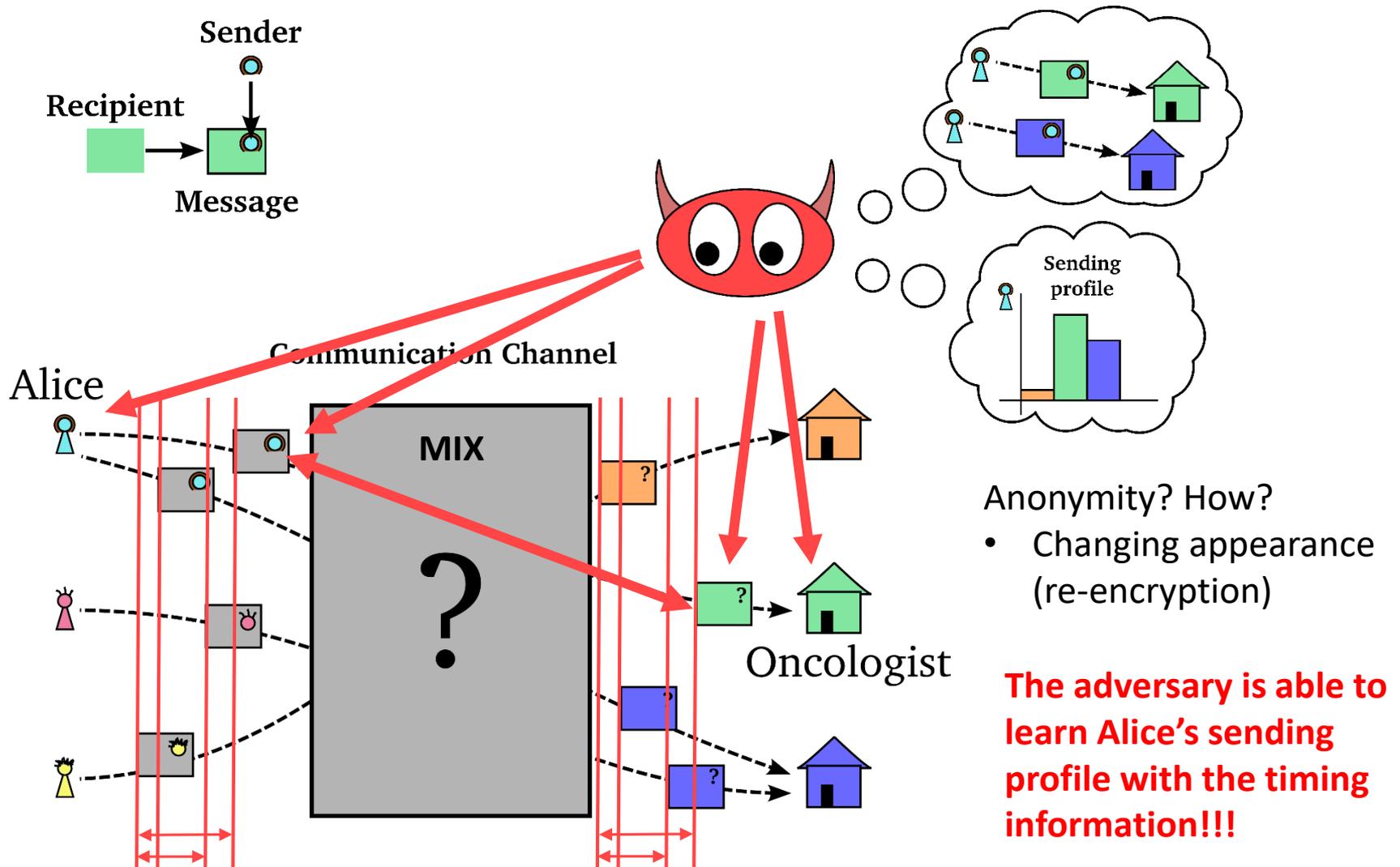
# Introduction. Mixes.



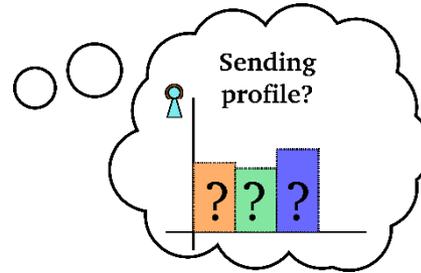
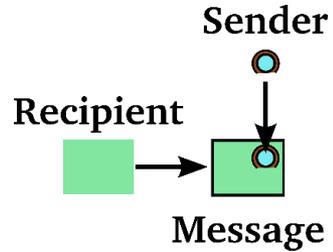
Anonymity? How?

- Changing appearance (re-encryption)

# Introduction. Mixes.

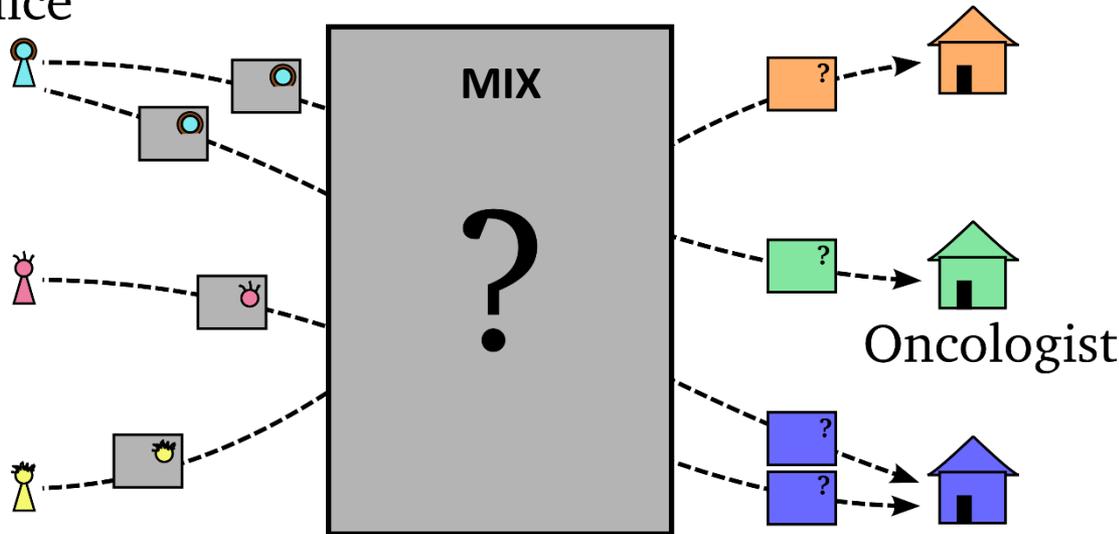


# Introduction. Mixes.



## ANONYMOUS Communication Channel

Alice



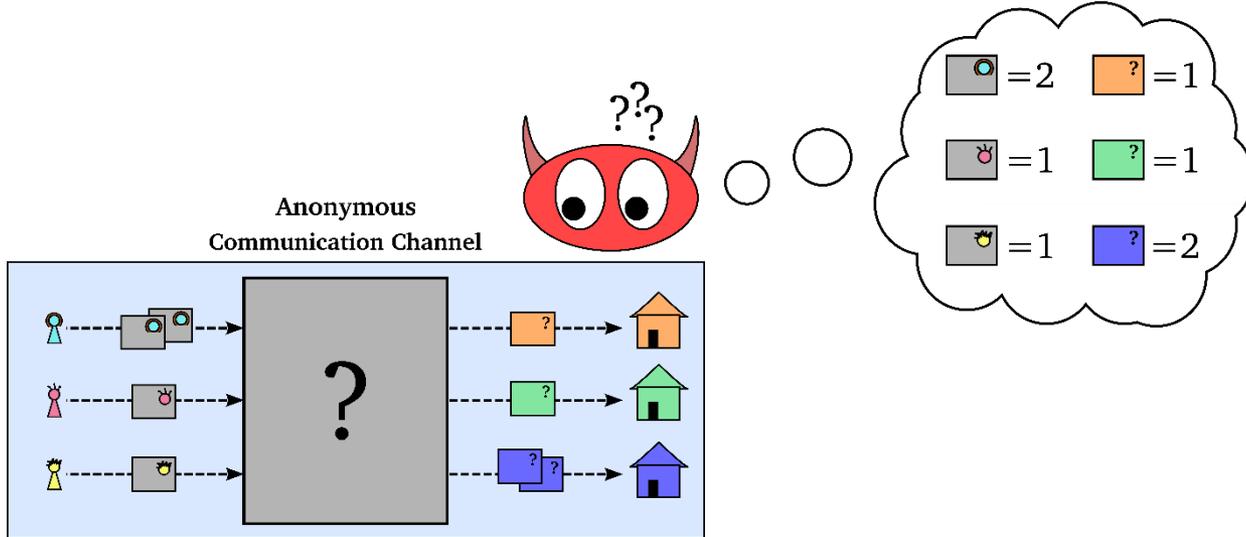
Anonymity? How?

- Changing appearance (re-encryption)
- Removing timing information (delays)

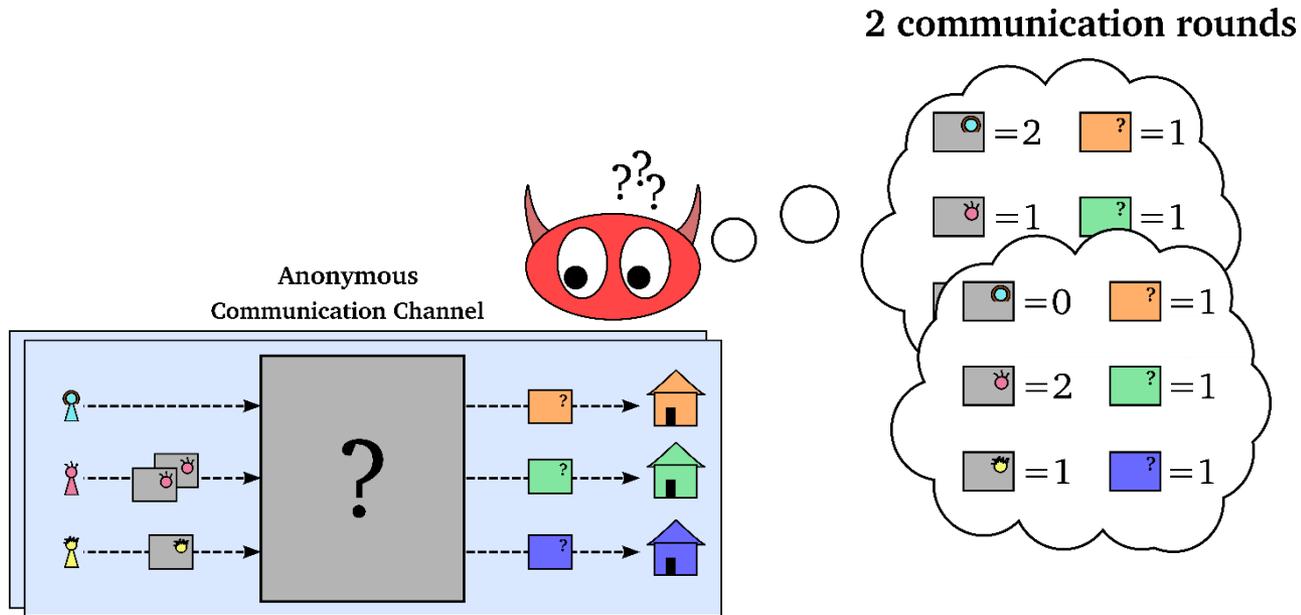


# Introduction. LSDA (I)

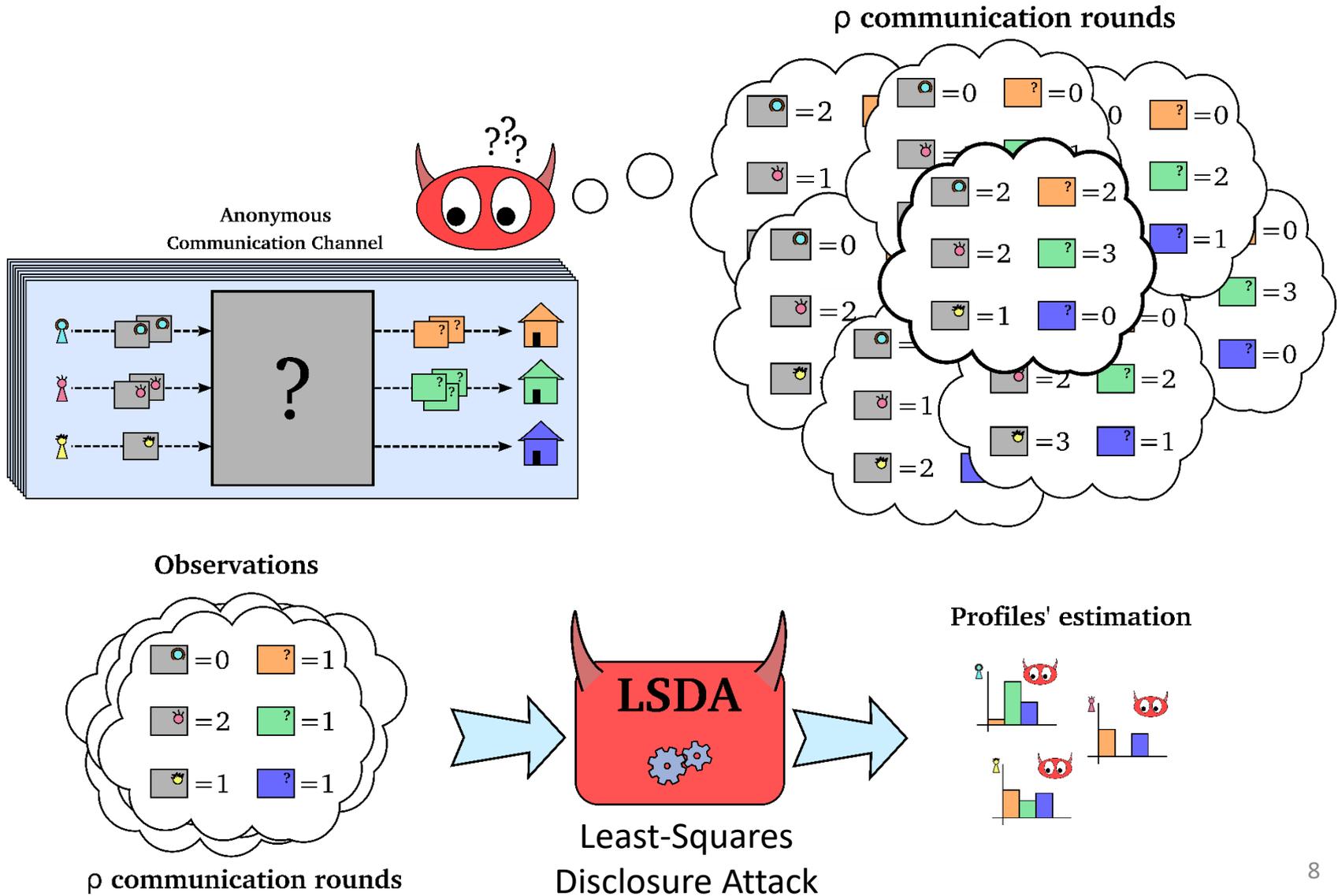
1 communication round



# Introduction. LSDA (I)

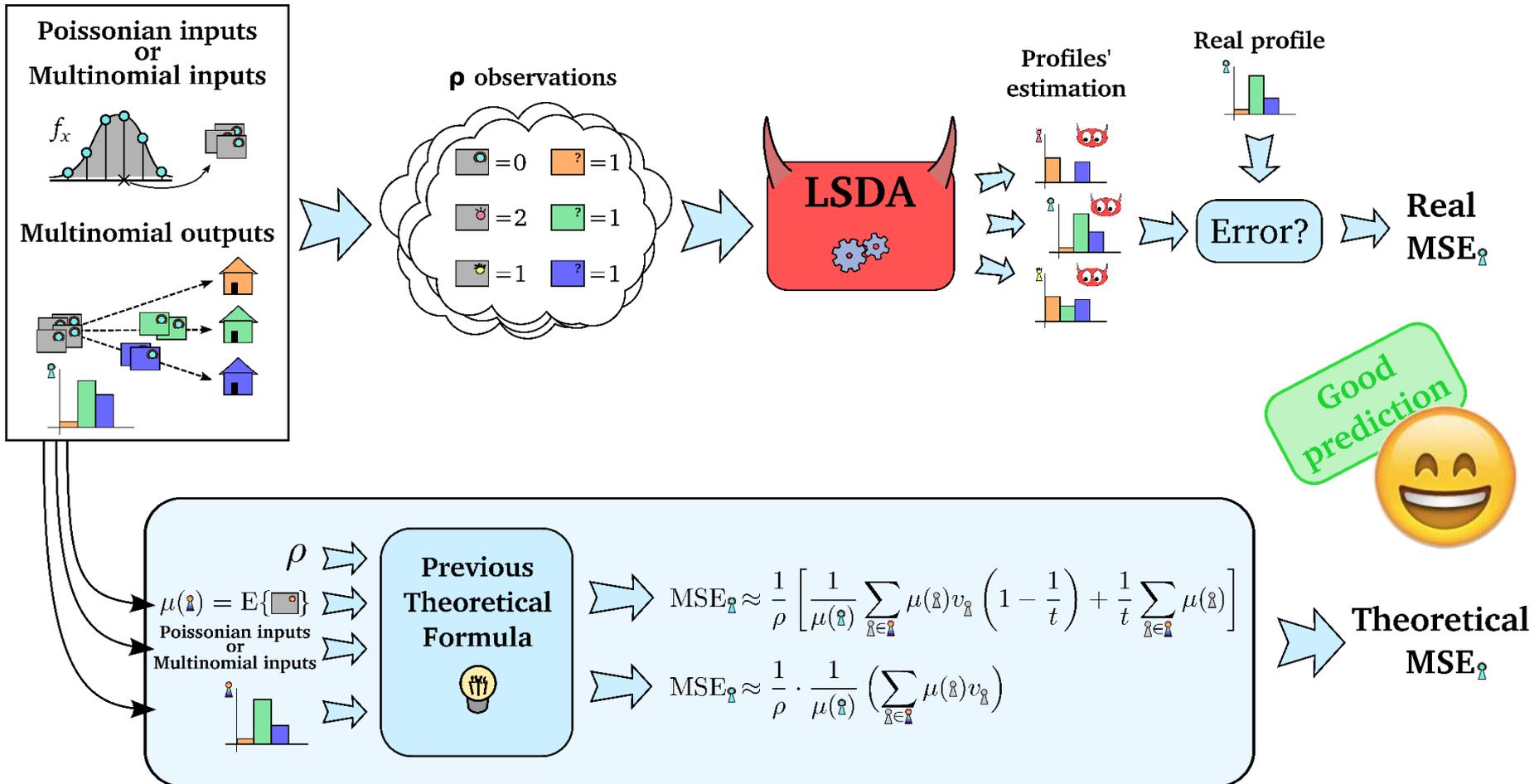


# Introduction. LSDA (I)



# Introduction. LSDA (II)

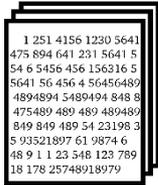
## SYSTEM MODEL



# Real data:

---

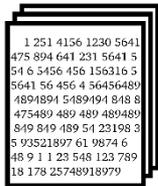
## Emails



220.000 emails sent between employees of *Enron Corporation*.

<http://www.cs.cmu.edu/~./enron/>

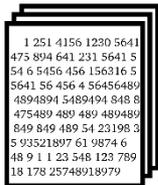
## Location



400.000 location check-ins from *Gowalla* social networking website.

<http://snap.stanford.edu/data/loc-gowalla.html>

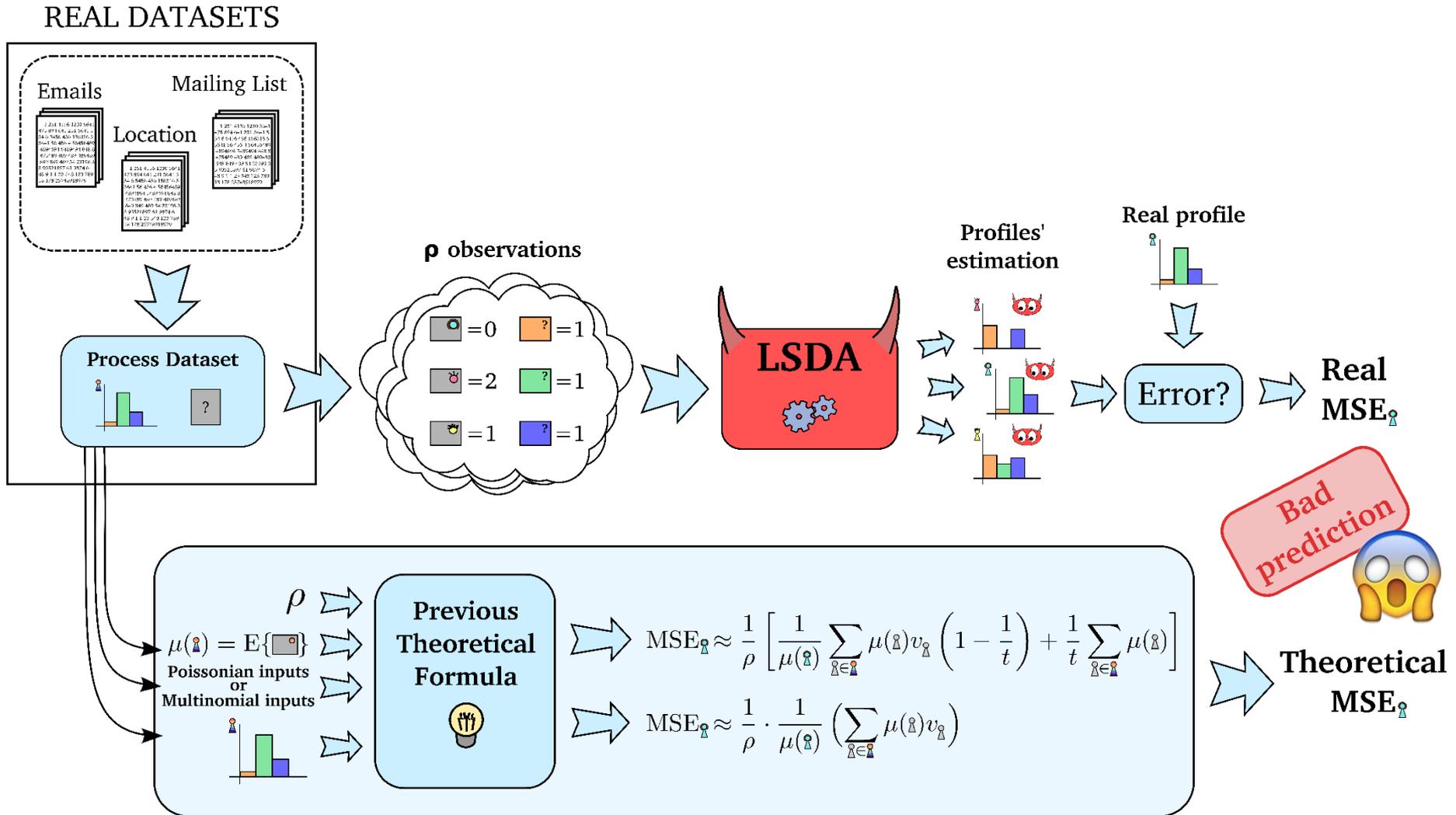
## Mailing List



180.000 posts to the public mailing lists of *Indimedia*

<http://lists.indimedia.org/>

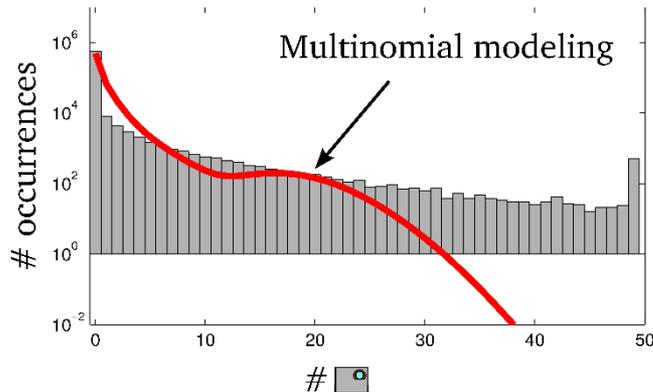
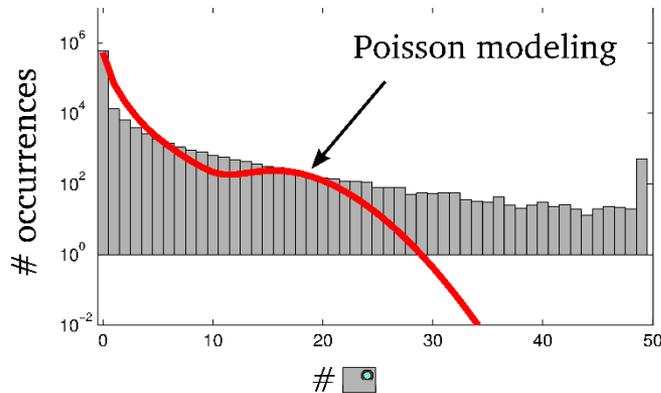
# Motivation: LSDA's analysis falls short



# Real-World Behavior. Input process.

BEFORE

Example: Emails dataset



NOW

We only need certain moments!



$$\mu(\text{[person icon]}) = E\{\text{[email icon]}\}$$

$$\mu_2(\text{[person icon]}) = E\{(\text{[email icon]} - E\{\text{[email icon]}\})^2\}$$

$$\mu_3(\text{[person icon]}) = E\{(\text{[email icon]} - E\{\text{[email icon]}\})^3\}$$

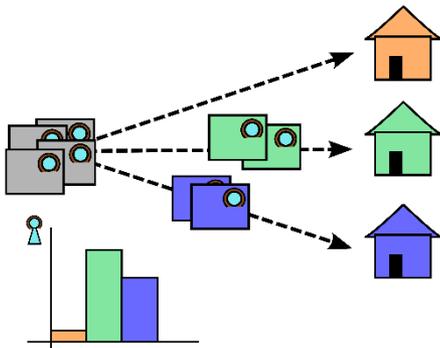
$$\mu_4(\text{[person icon]}) = E\{(\text{[email icon]} - E\{\text{[email icon]}\})^4\}$$

# Real-World Behavior. Output process.

Average number of receivers chosen

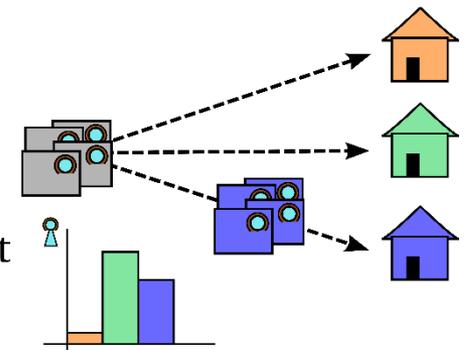
**BEFORE**

Multinomial choice of receivers



**NOW**

Also, maximum variance



#messages	Email	Location	Mailing List
=2	1.85	1.03	1.29
=3	2.71	1.05	1.46
=4	3.53	1.06	1.53
=5	4.40	1.08	1.53
≥6	13.56	1.11	1.57

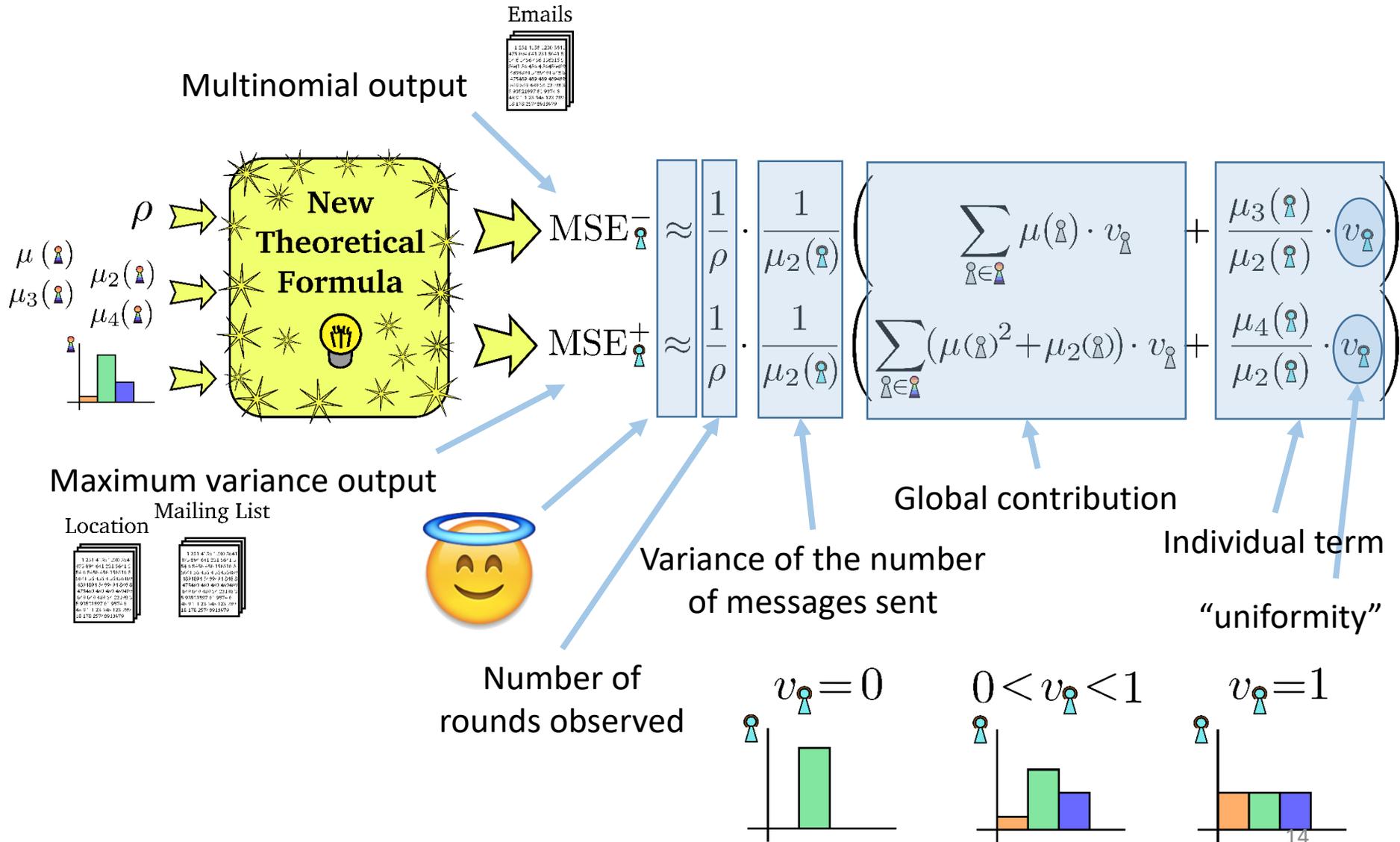
Emails



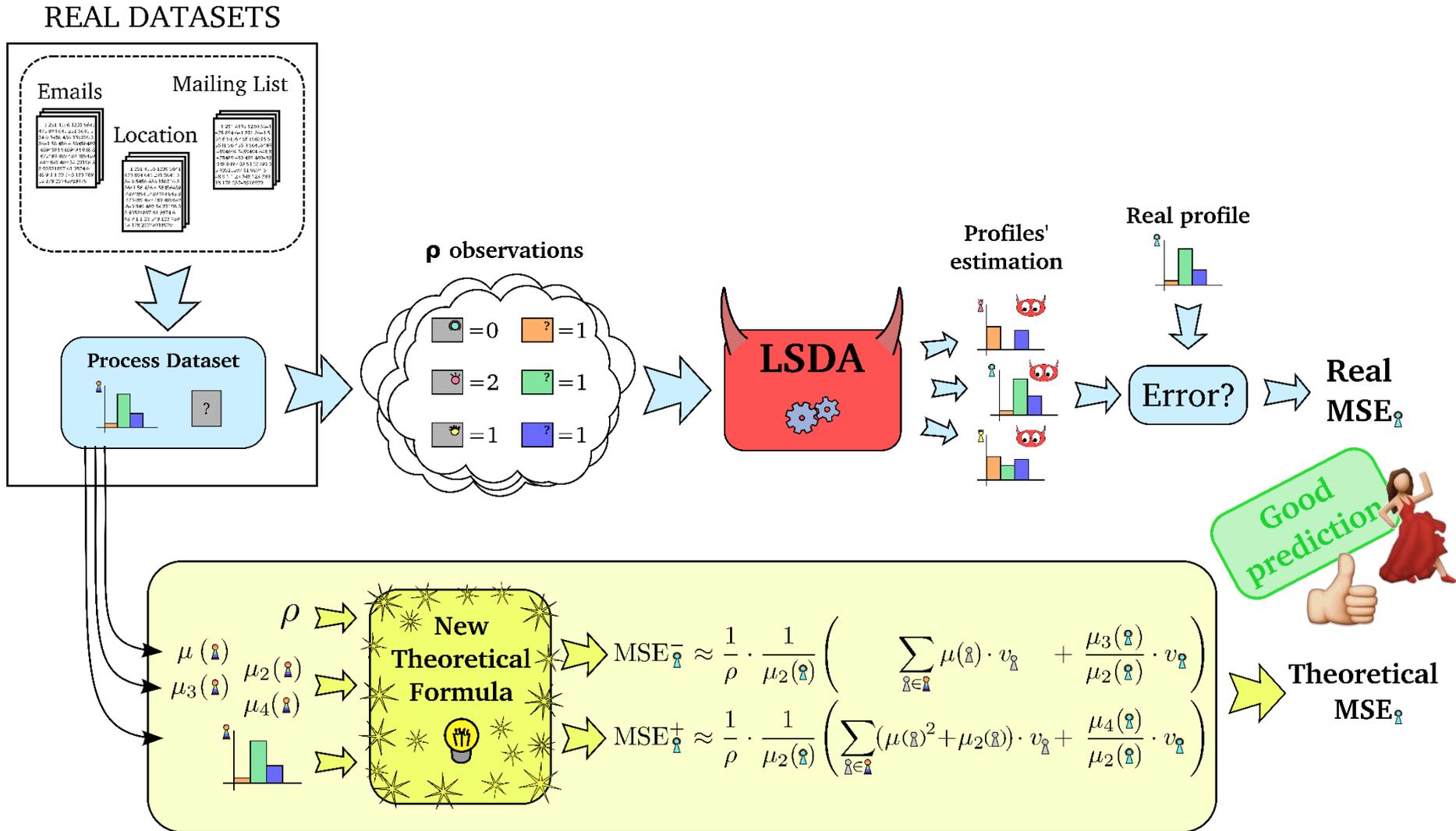
Mailing List  
Location



# New Theoretical Analysis (I)

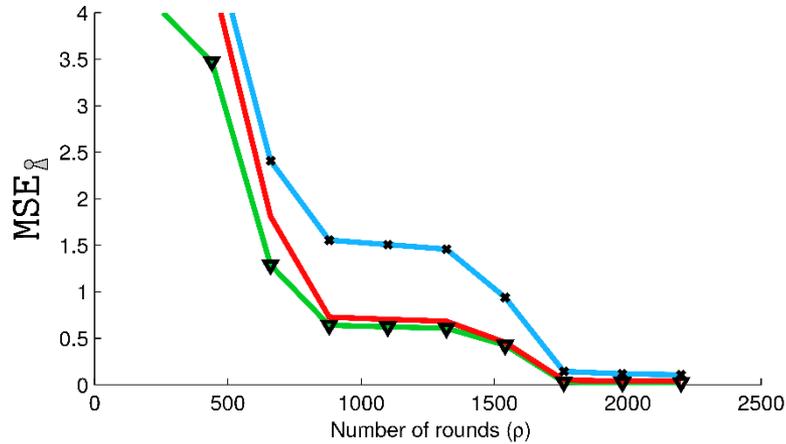


# New Theoretical Analysis (II)

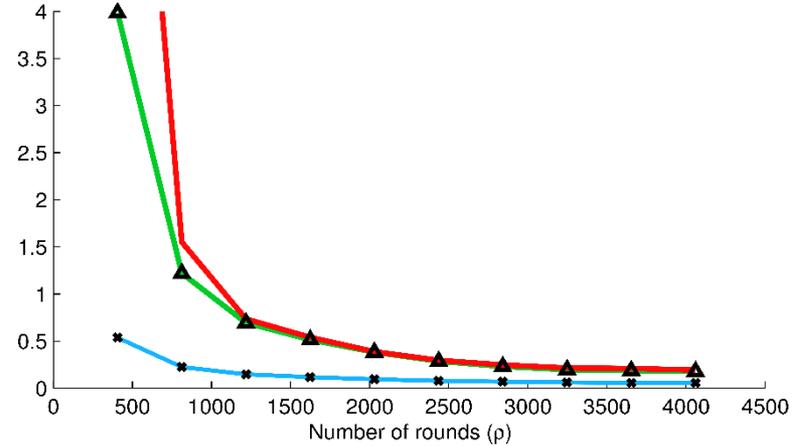


# Results

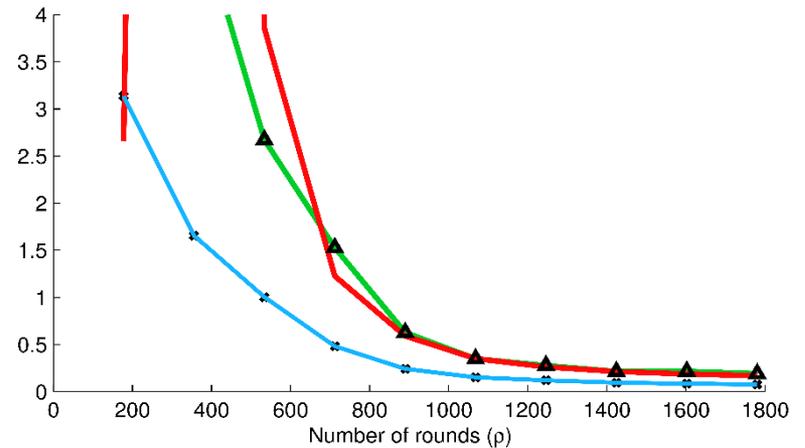
## Emails



## Location



## Mailing List



- Real MSE
- \* Previous Theoretical MSE
- ▼ New MSE<sup>-</sup> (multinomial)
- ▲ New MSE<sup>+</sup> (maximum variance)

**Thank you!!**

simonoya@gts.uvigo.es

