



**COSIC**

8<sup>th</sup> Privacy Enhancing Technologies Symposium  
PETS'08

# Perfect Matching Disclosure Attacks

**Carmela Troncoso**

Bart Preneel

Benedikt Gierlichs

Ingrid Verbauwhede

KULeuven COSIC/ESAT (Belgium)

23<sup>rd</sup> July Leuven Belgium

# Anonymous Communications

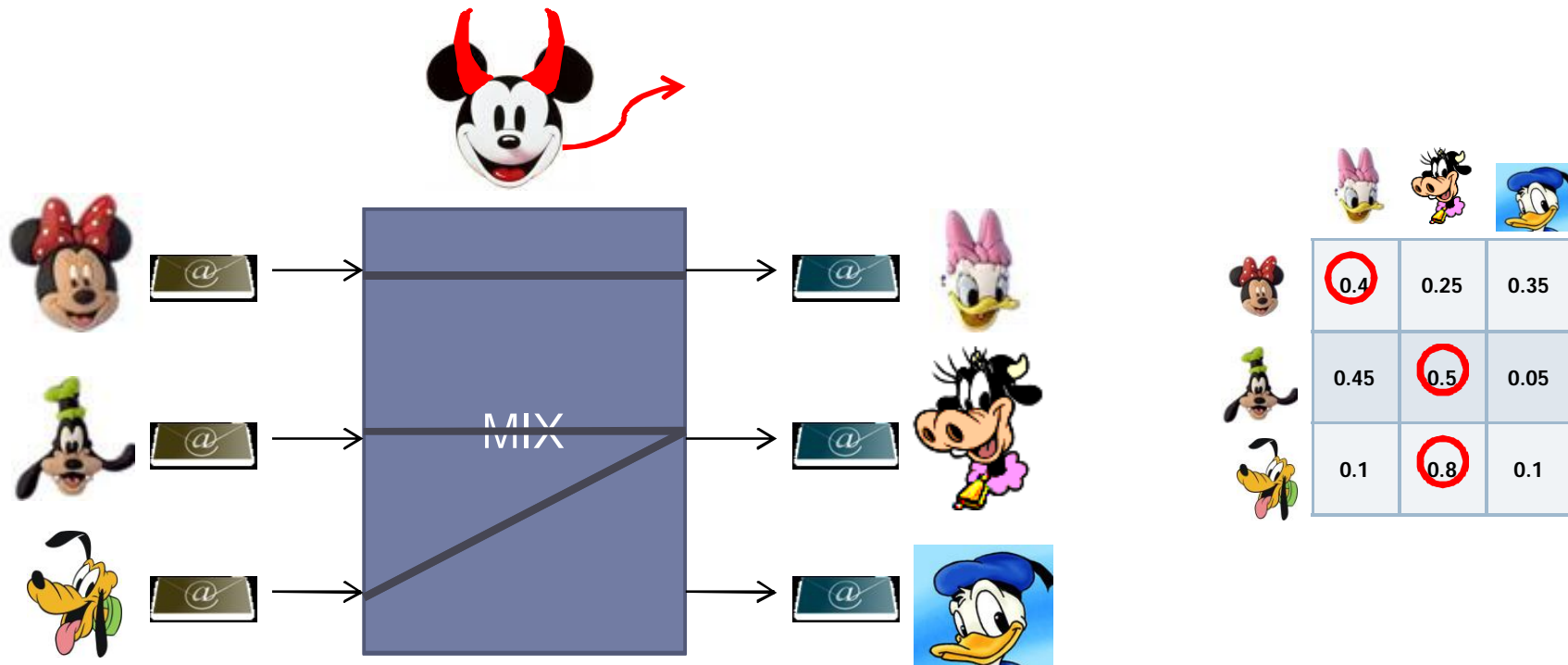
- ▶ "Tell me who your friends are. ." => Anonymous communications to hide communication partners
- ▶ High latency systems (e.g. anonymous remailers) use mixes [Chaum 81]: hide input/output relationship



- ▶ Disclosure attacks: exploit patterns to uncover links
  - ▶ Global passive attacker
  - ▶ Simple model: restrictive assumptions on user behavior
  - ▶ Exact solution [Kes03] = NP-problem
  - ▶ Statistical Disclosure Attacks (SDA) [Dan03]

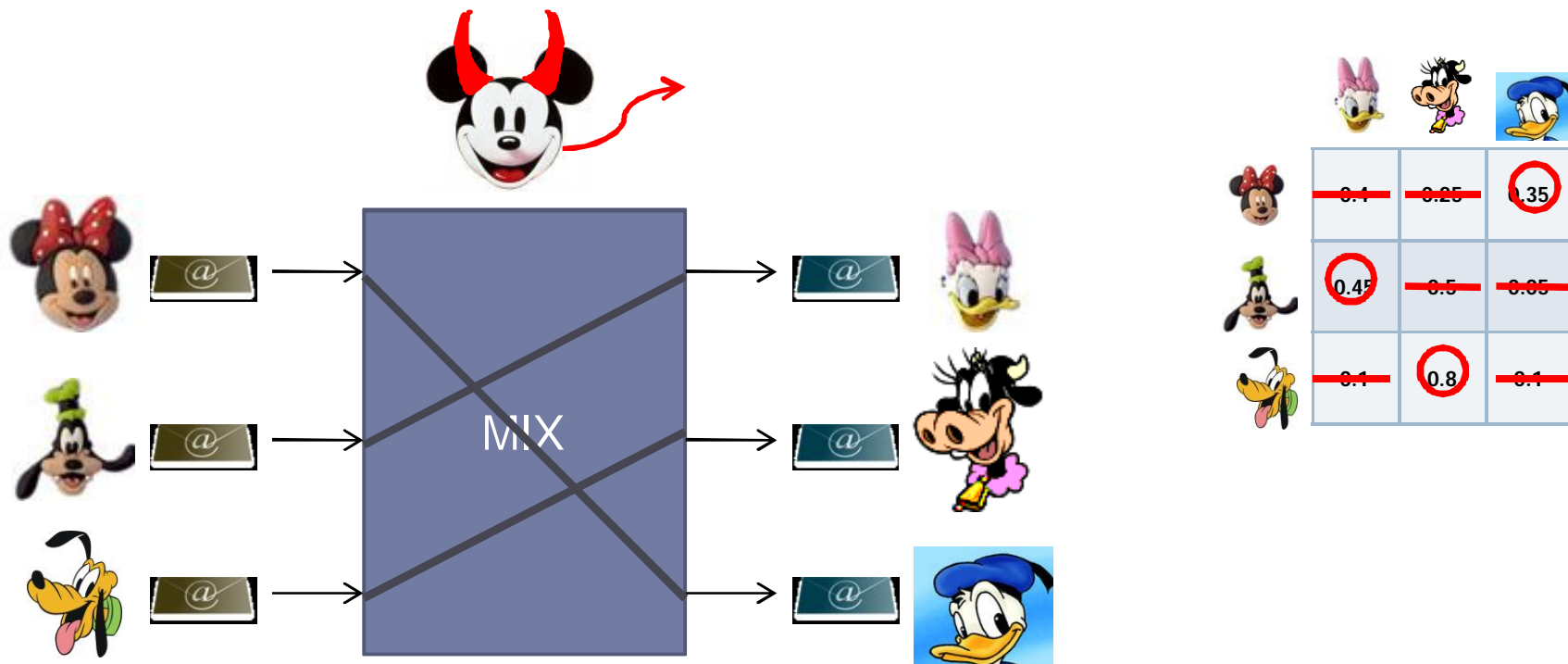
# Intuition behind our attack

- ▶ Who communicates with whom?
- ▶ Previous work: users treated independently [Dan03,DDT07]
  - ▶ Take the most likely receiver for each of the users



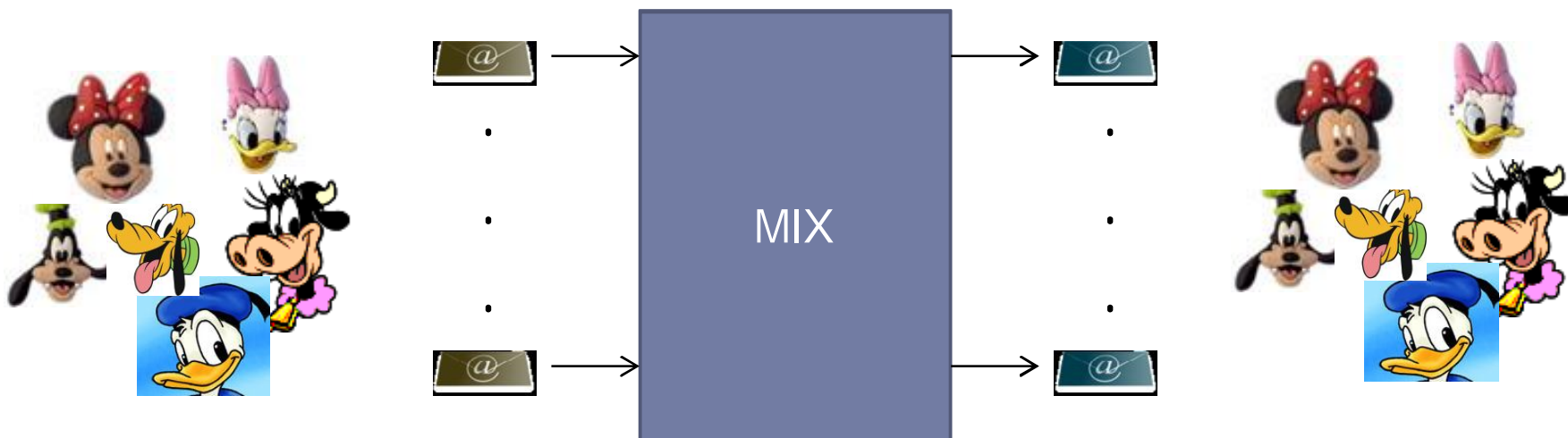
# Intuition behind our attack

- ▶ Who communicates with whom?
- ▶ Previous work: users treated independently [Dan03,DDT07]
- ▶ Why don't we use all information available?
  - ▶ If Pluto sends to Clarabella, Goofy cannot send to Clarabella



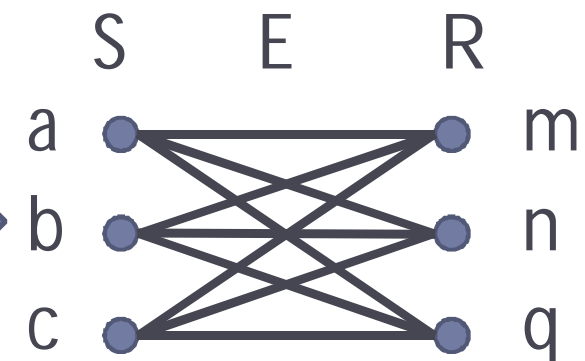
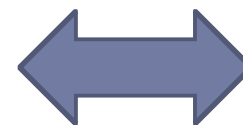
# System model

- ▶ Threshold mix with threshold  $t$
- ▶ Users send independently
- ▶  $P_x$  denotes the profile of user  $x$
- ▶ Friendship: " $y$  is friend of  $x$  if  $x$  sends a message to  $y$  with non-zero probability  $P_x(y)$ "



# How to do it?

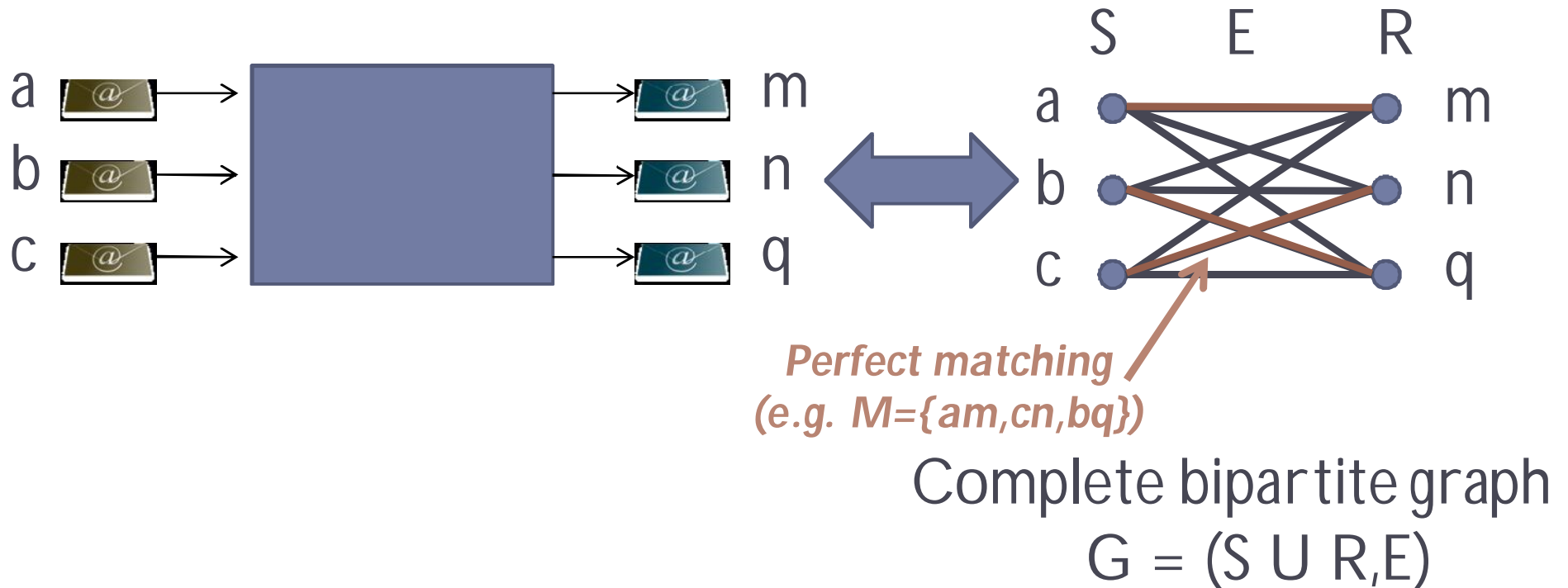
## ► Graph theory



Complete bipartite graph  
 $G = (S \cup R, E)$

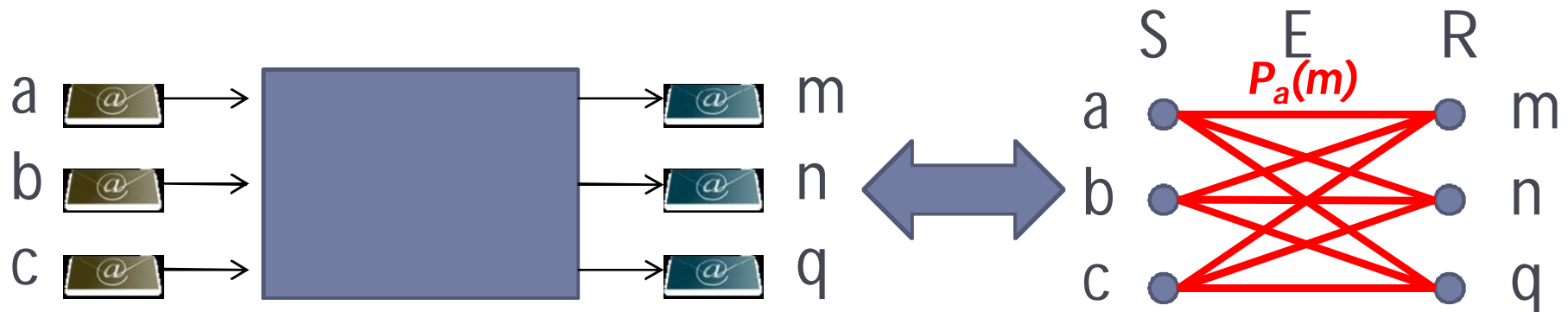
# How to do it?

## ► Graph theory



# How to do it?

## ▶ Graph theory

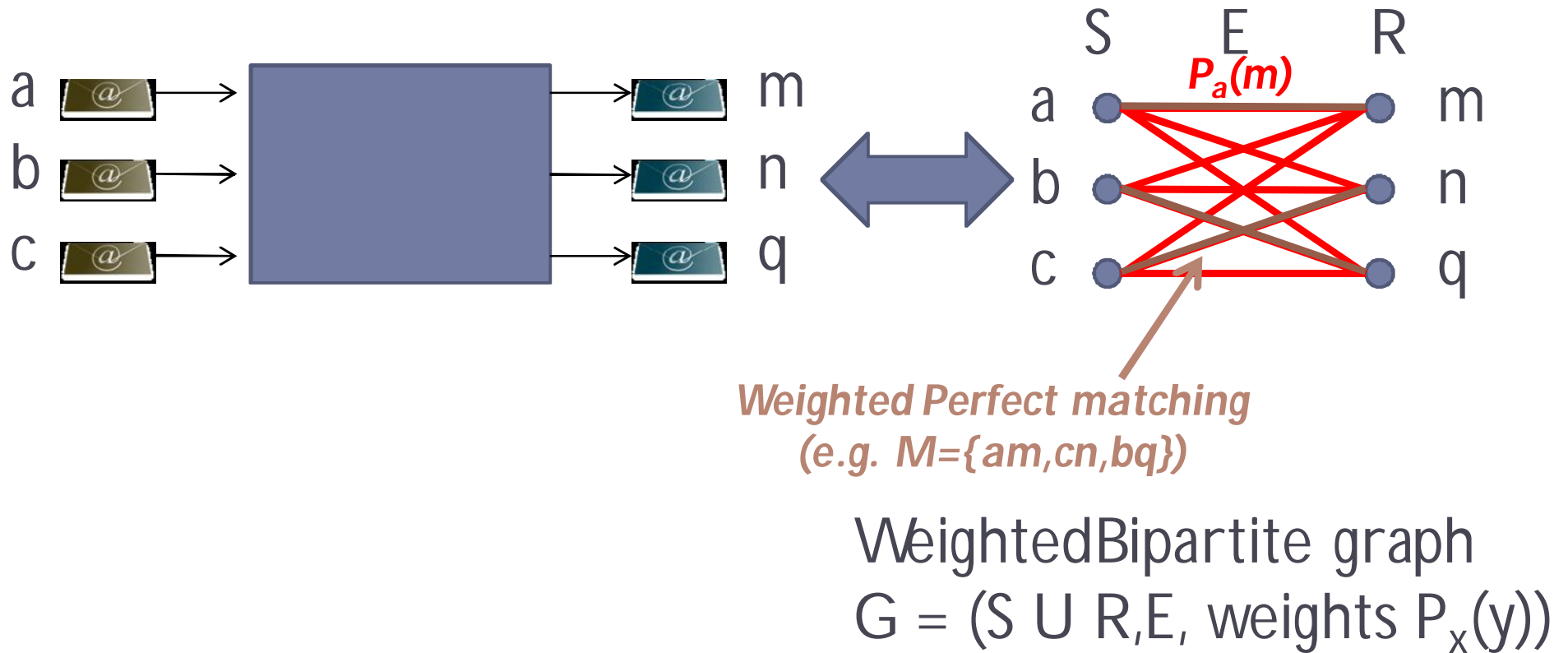


WeightedBipartite graph  
 $G = (S \cup R, E, \text{weights } P_x(y))$



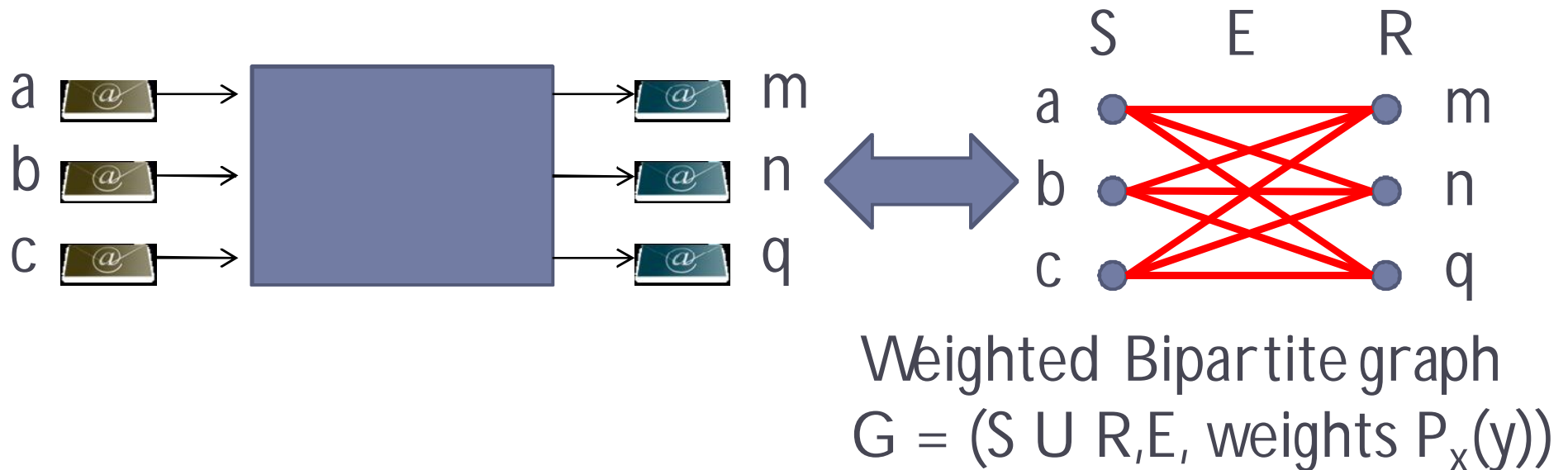
# How to do it?

## ▶ Graph theory



# How to do it?

## ▶ Graph theory



## ▶ Optimization problem

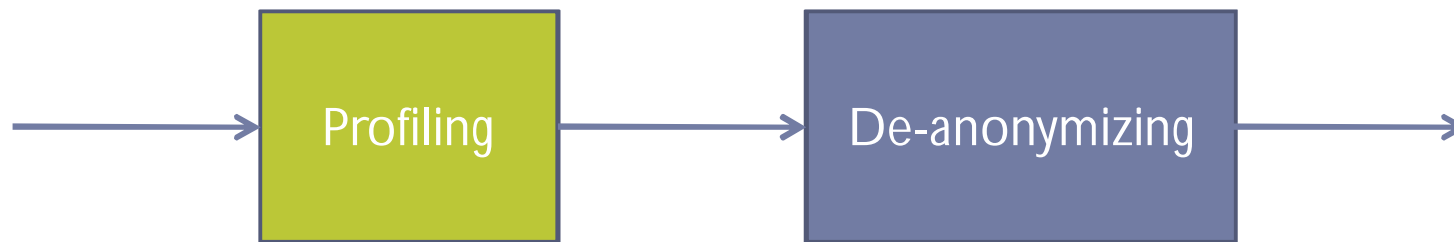
$$\max(p(M | S, R)) \Leftrightarrow \max(p(M)) \quad (\text{from Bayes})$$

$$\max(p(M)) = \max\left(\prod_{xy \in M} p_x(y)\right) \Leftrightarrow \max\left(\sum_{xy \in M} \log(p_x(y))\right)$$

**Maximum weighted perfect matching**  
**Efficient solution: linear assignment problem**

# The Attack: profiling users

---



# The Attack: profiling users

- ▶ Observe the system during  $T$  rounds collecting  $S$  and  $R$  in each of them
- ▶ Statistical Disclosure Attack (SDA) finds the likely set of friends of each user  $P_{x, SDA}$

$$O = \frac{1}{t} P_{Alice} + \frac{1}{t} P_x + \dots + \frac{1}{t} P_w$$

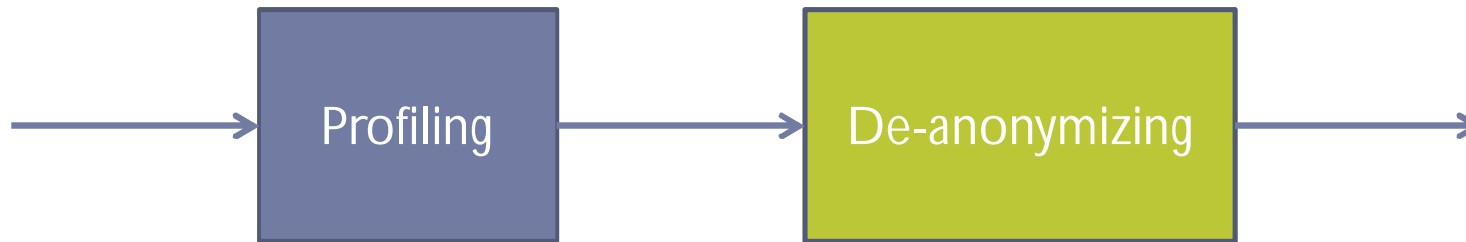
- ▶ Only Alice has friends, the rest of users send uniformly

$$O = \frac{1}{t} P_{Alice} + \frac{t-1}{t} P_x$$

$$\tilde{P}_{Alice} \approx \sum_{i=1}^T O_i - (t-1) P_x$$

# The Attack: de-anonymizing users

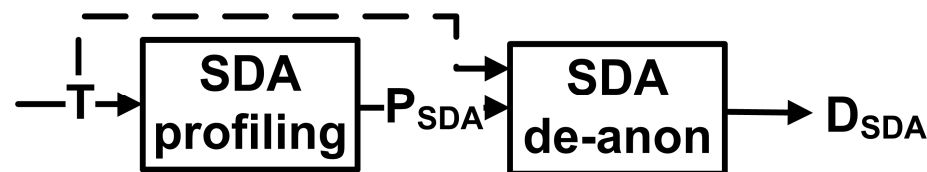
---



# The Attack: de-anonymizing users




## Statistical Disclosure Attack (SDA)




- Given  $P_{x,SDA}$  chooses the most likely receiver independently

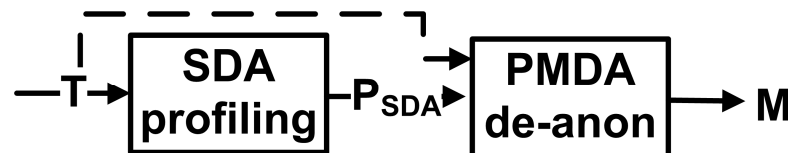
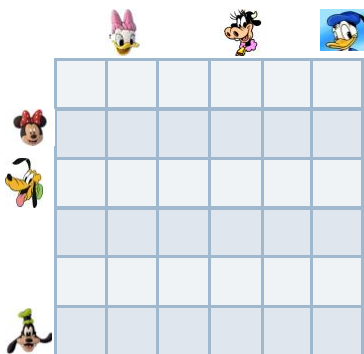








## Perfect Matching Disclosure Attack (PMDA)


- Considers all the users in the round simultaneously











|   |      |      |      |
|---|------|------|------|
|  | 0.4  | 0.25 | 0.35 |
|  | 0.45 | 0.5  | 0.05 |
|  | 0.1  | 0.8  | 0.1  |

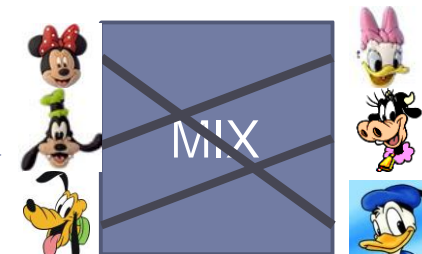
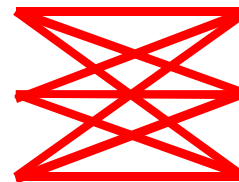



|   |   |   |
|---|---|---|
|  |  |  |
|  |  |  |
|   |   |   |
|   |   |   |
|   |   |   |
|   |   |   |



|   |   |   |   |
|---|---|---|---|
|   |  |  |  |
|  | 0.4   | 0.25  | 0.35  |
|  | 0.45  | 0.5   | 0.05  |
|  | 0.1   | 0.8   | 0.1   |

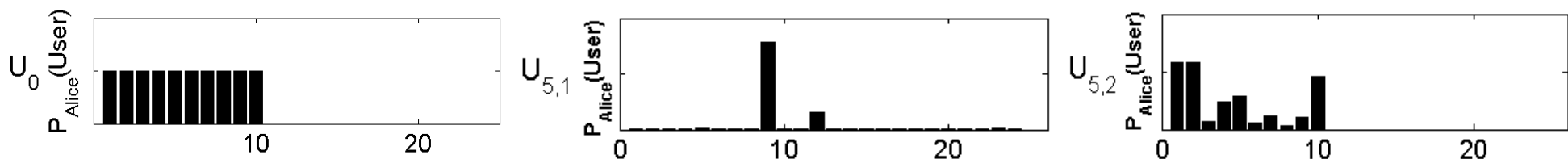
$P'$



# Evaluation – SDA vs PMDA

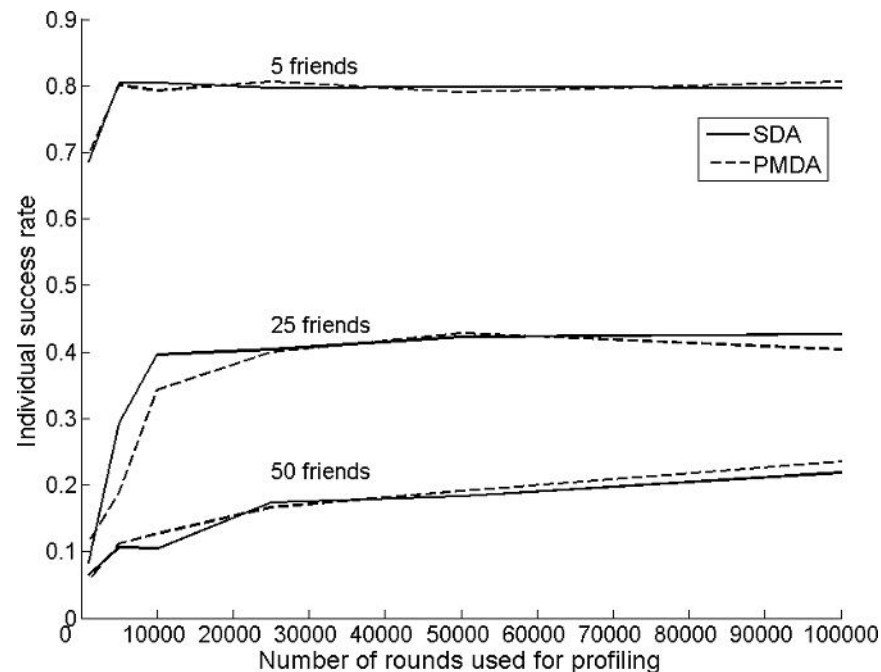
|                      |                             |
|----------------------|-----------------------------|
| Size of population   | 1000                        |
| Sending rate         | $\lambda$                   |
| Threshold            | 100                         |
| Rounds for profiling | 1k,5k,10k,25k,50k,100k      |
| Rounds de-anonymized | 5k (~500 messages per user) |

- Two populations
  - $U_0$ : only Alice has a fixed number  $r$  friends amongst which she chooses at random [almost Kes03, Dan03]
  - $U_5$ : every user has a random number of friends amongst which they choose with non-uniform probability



# Results $U_0$

- ▶ Only Alice's results (5, 25 and 50 friends)

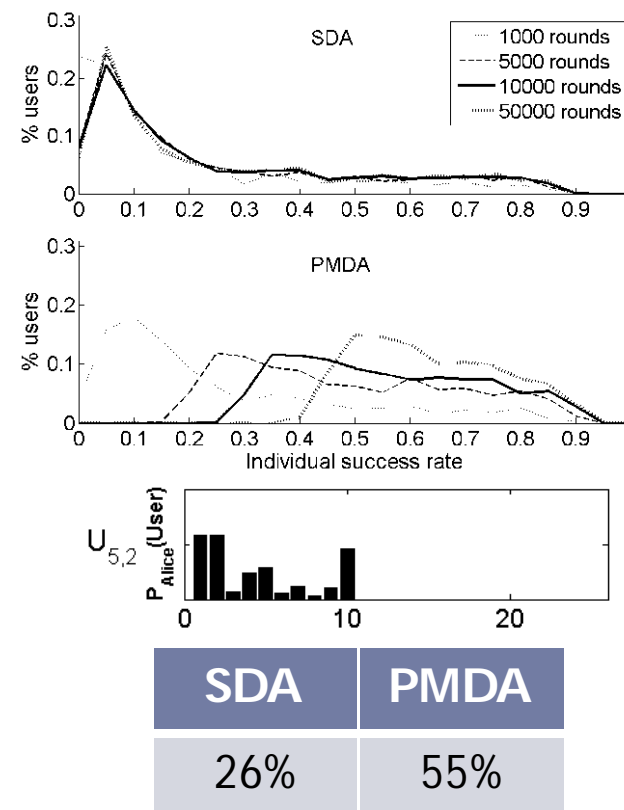
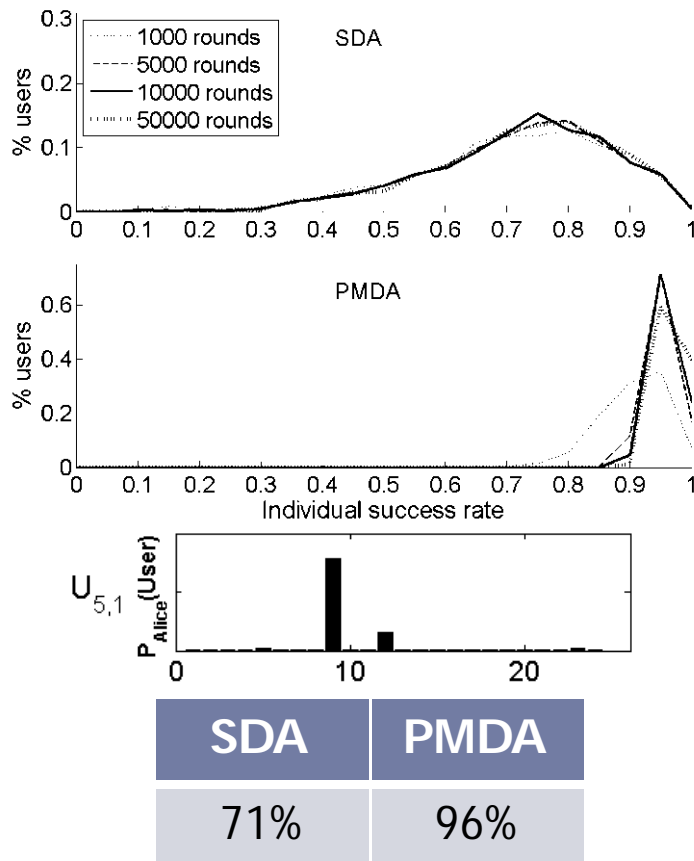


- ▶ No extra information
- ▶ Alice chooses uniformly (better small number of friends)
- ▶ More rounds, more accuracy



# Results $U_5$

- **Individual success rate** : accuracy in de-anonymizing messages from one sender



# Scalability

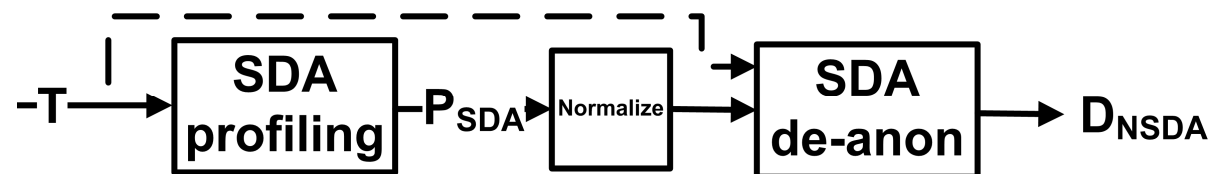
- ▶ Timings de-anonymizing 5000 rounds with p rofiles constructed after 50000 rounds observed

| Attack        | t=100   |                         | t=500    | t=1000   |
|---------------|---------|-------------------------|----------|----------|
|               | Time    | Success rate mean (min) | Time     | Time     |
| SDA profiling | 3min    | -                       | 38.33min | 66.16min |
| SDA de-anon   | 10min   | 25.6%(0.0%)             | 3.48h    | 12.9h    |
| PMDA de-anon  | 10.2min | 62.9%(38.8%)            | 12.9h    | 4.69days |

- ▶ Regular PCs
- ▶ Non-optimized (high level interpreted language)
- ▶ Linear assignment problem can be parallelized

# Normalized SDA – Accuracy vs speed

- ▶ Normalized Statistical Disclosure Attack (NSDA)
  - ▶ SDA profiling + construction of  $P'$  + normalization + SDA de-anon

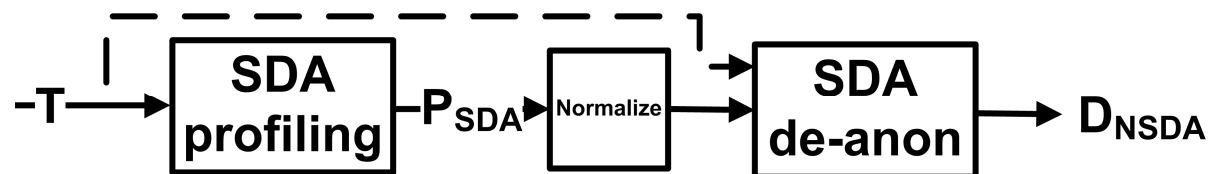


- ▶ Normalization: all rows and columns of  $P'$  add up to 1
  - ▶ iterative proportional fitting
  - ▶ spreads information of an element over the whole matrix
  - ▶ eliminates noise

$$P' = \begin{pmatrix} 0.4006 & 0.4208 & 0.1786 \\ 0.7810 & 0.1432 & 0.0757 \\ 0.0997 & 0.4580 & 0.4424 \end{pmatrix} \xrightarrow{\text{normalize}} \begin{pmatrix} 0.2776 & 0.4369 & 0.2856 \\ 0.6673 & 0.1834 & 0.1494 \\ 0.0552 & 0.3798 & 0.5651 \end{pmatrix}$$

# Normalized SDA – Accuracy vs speed

- ▶ Normalized Statistical Disclosure Attack (NSDA)
  - ▶ SDA profiling + construction of  $P'$  + normalization + SDA de-anon



- ▶ Normalization: all rows and columns of  $P'$  add up to 1
  - ▶ iterative proportional fitting
  - ▶ spreads information of an element over the whole matrix
  - ▶ eliminates noise

| Attack       | t=100    |              | t=500 | t=1000   |
|--------------|----------|--------------|-------|----------|
|              | Time     | Success rate | Time  | Time     |
| SDA de-anon  | 10min    | 25.6%(0.0%)  | 3.48h | 12.9h    |
| PMDA de-anon | 10.2min  | 62.9(38.8%)  | 12.9h | 4.69days |
| NSDA de-anon | 13.33min | 60.2(33.5%)  | 4.28h | 15.3h    |

# Enhanced profiling

- ▶ SDA considers all receivers in a round as equally likely

$$O = \frac{1}{t} P_{Alice} + \frac{1}{t} P_x + \dots + \frac{1}{t} P_w$$

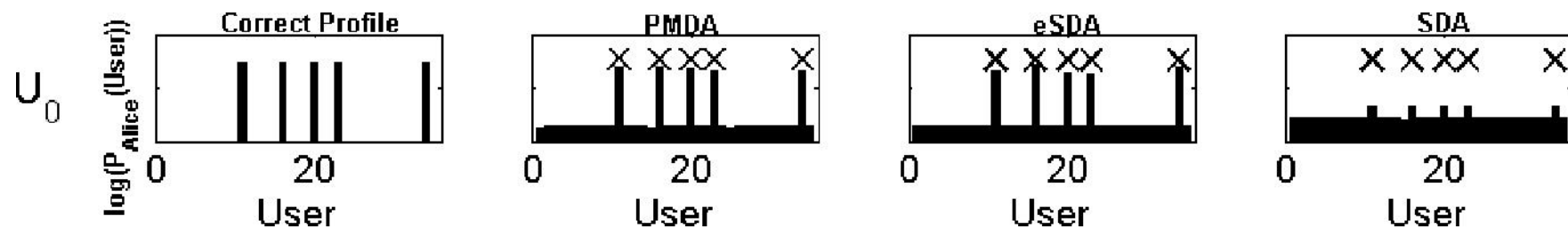
- ▶ We can use the matchings obtained from PMDA de-anonymization to improve this result yielding  $P_{PMDA}$ 
  - ▶ Assign  $z$  to the receiver assigned by PMDA
  - ▶ Assign  $(1-z)/(t-1)$  to the rest
    - ▶  $z=(1-z)/(t-1)$  is the SDA and  $z < (1-z)/(t-1)$  hides actual relationships

$$O = z P_{Alice} + \frac{1-z}{t-1} P_x + \dots + \frac{1-z}{t-1} P_w$$

- ▶ We can apply the same philosophy to SDA de-anonymization obtaining  $P_{eSDA}$

# Enhanced profiling

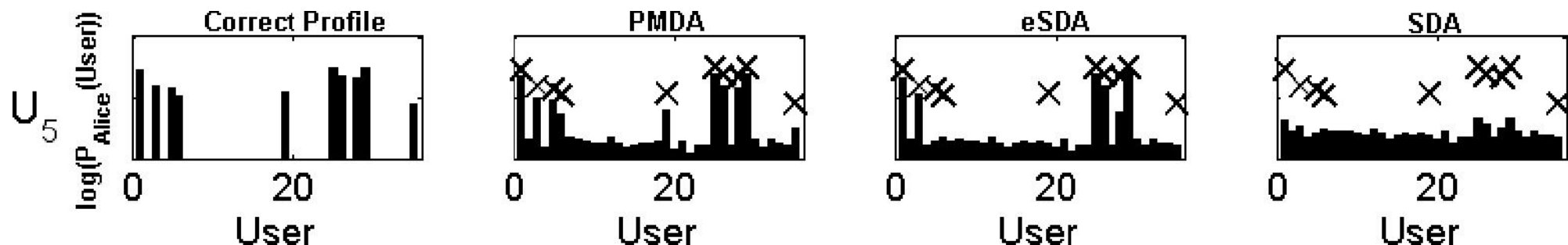
- ▶  $U_0$  population, 5 friends



- ▶ All methods can distinguish friends (even if the number of friends is unknown)
- ▶ Enhanced methods increase contrast

# Enhanced profiling

## ► $U_5$ population



- SDA cannot distinguish friends
- eSDA only detects the 'best' friends
- PMDA all friends have higher probability but threshold unclear

# Conclusions

---

- ▶  $U_5$  more generic user model
- ▶ Perfect Matching DisclosureAttack
  - ▶ Considers all users in a round simultaneously
  - ▶ More accurate than previous methods without assumptions on the underlying user behaviour
- ▶ Normalized Statistical DisclosureAttack
  - ▶ Less accurate but faster
- ▶ Enhanced Profiling Methodologies



# Future lines

---

- ▶ Further generalization of the user behaviour
  - ▶ Sendingrate
  - ▶ Behaviourvariance over time
- ▶ Extension to pool mixes
- ▶ Improve efficiency of PMDA
  - ▶ Parallelize attack
  - ▶ Parallelize LinearAssignment Problem solver

---



Thanks for your  
attention

Carmela.Troncoso@esat.kuleuven .be