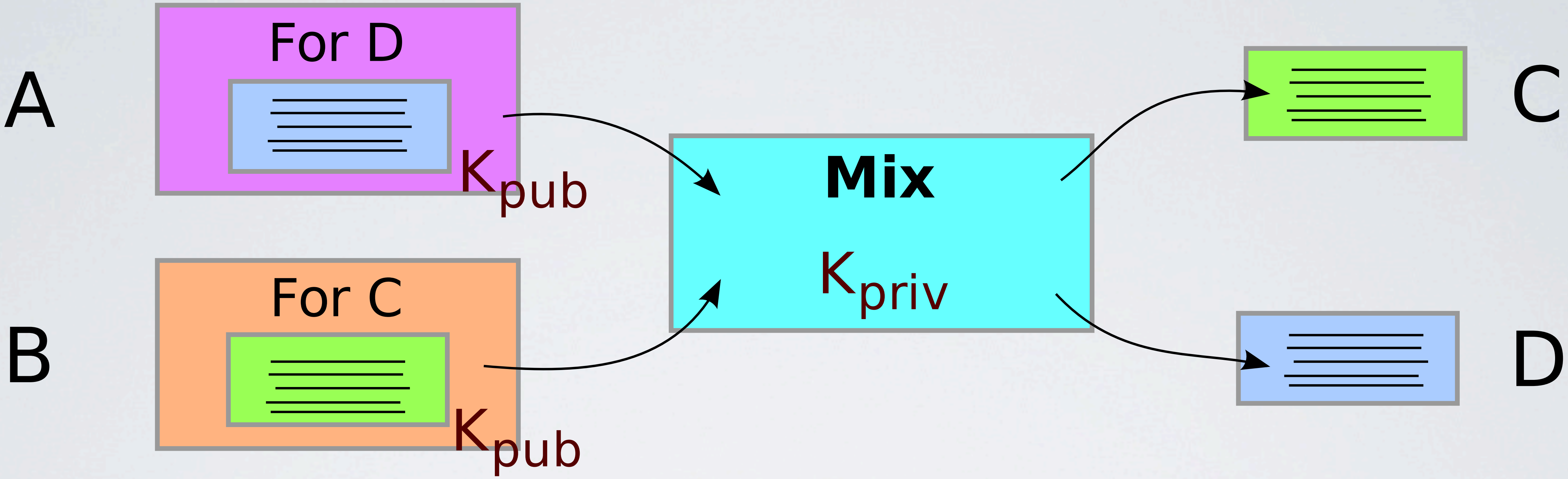
QUANTIFYING AND MEASURING ANONYMITY

Steven J. Murdoch
University of Cambridge

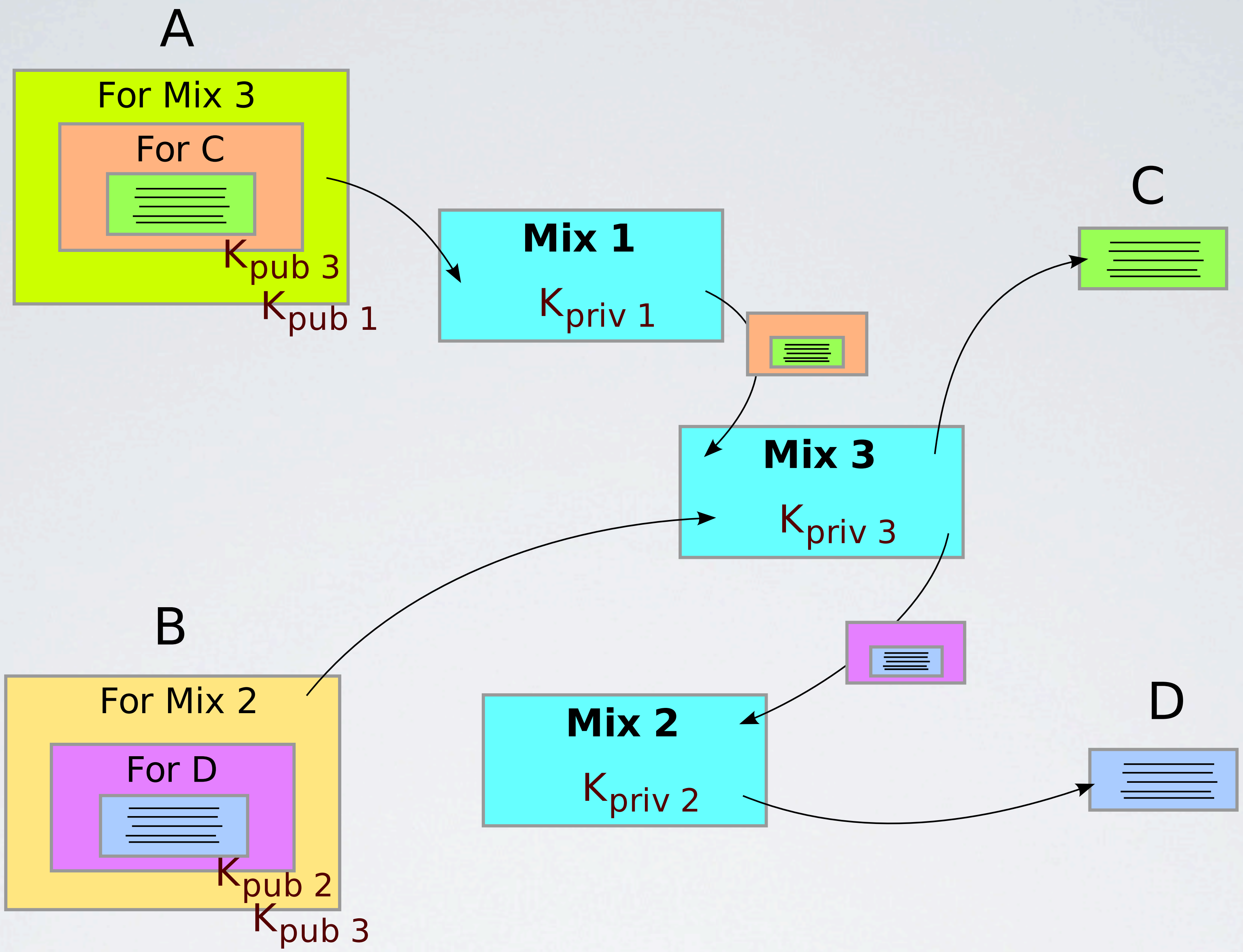
WHY IS ANONYMITY INTERESTING?

- Metrics for cryptography often don’t matter
  - Key length of 192 bits or 256 bits is irrelevant
- Anonymity can never reach the same levels of security
  - At best you are 1 in 7 billion people (33-bit key length)
- Exact level of security much more importance when there is no safety margin



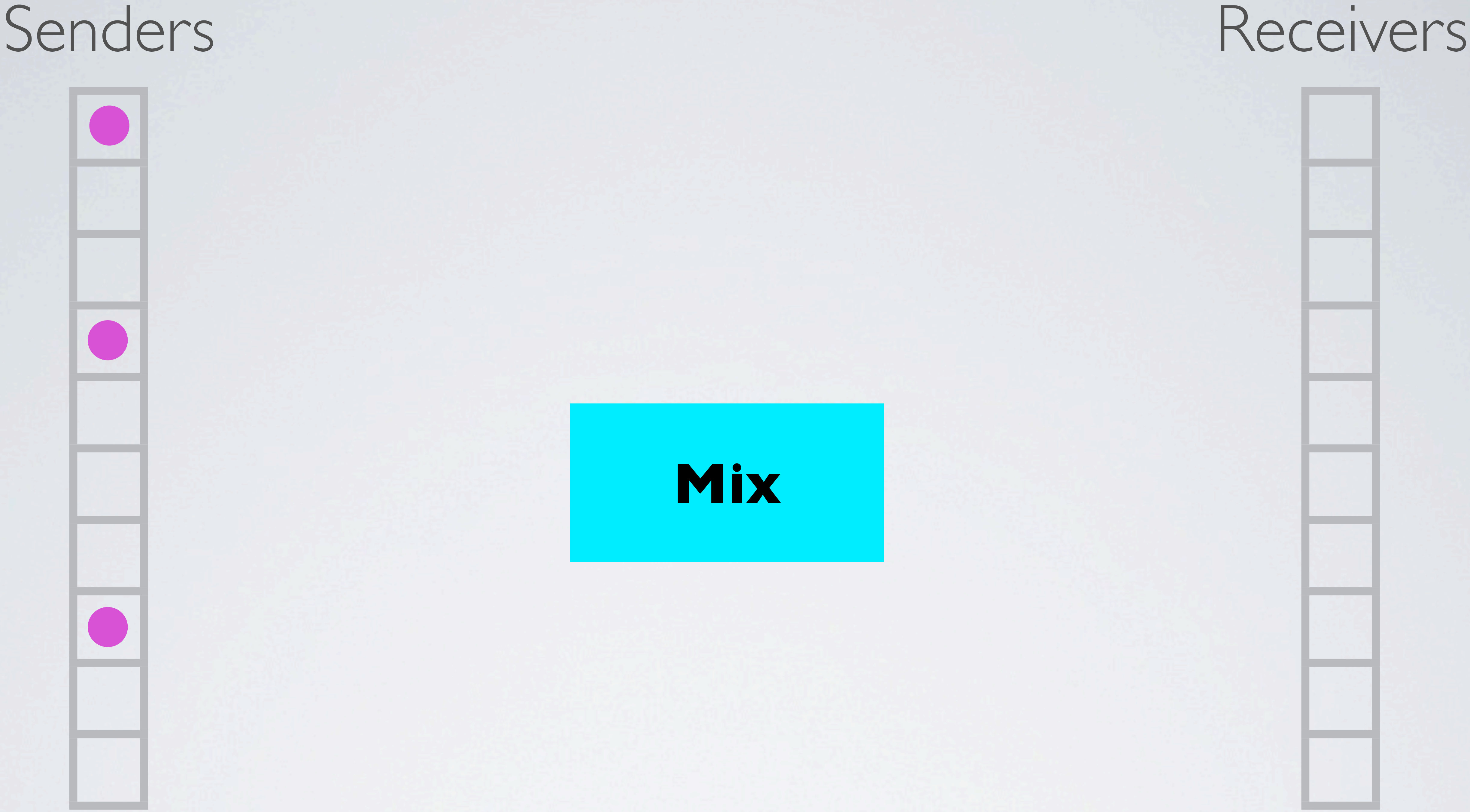
ONE-HOP MIX

If mix compromised, system insecure



MULTI-HOP MIX

Damage of single-mix compromise reduced



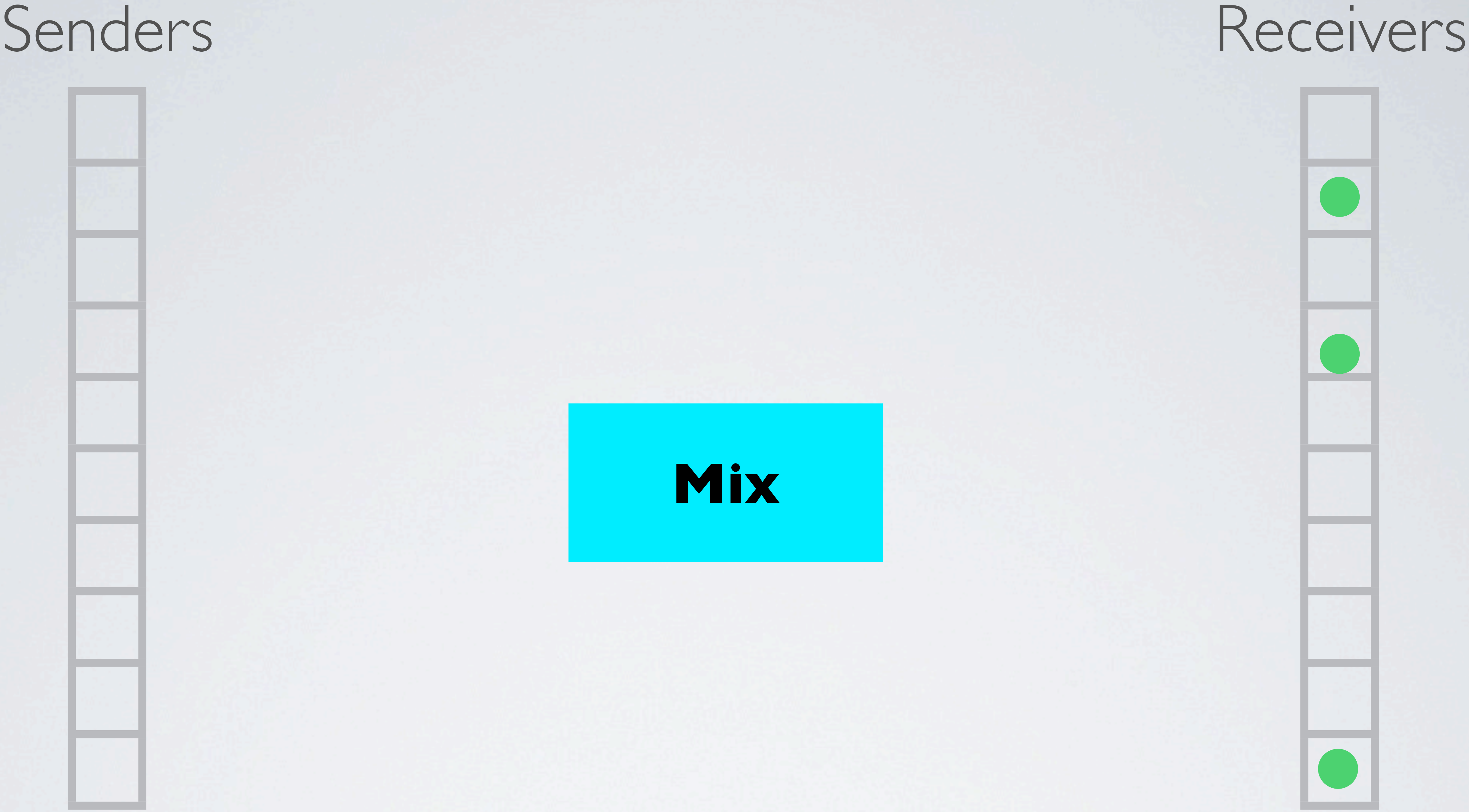
Senders

Mix

Receivers

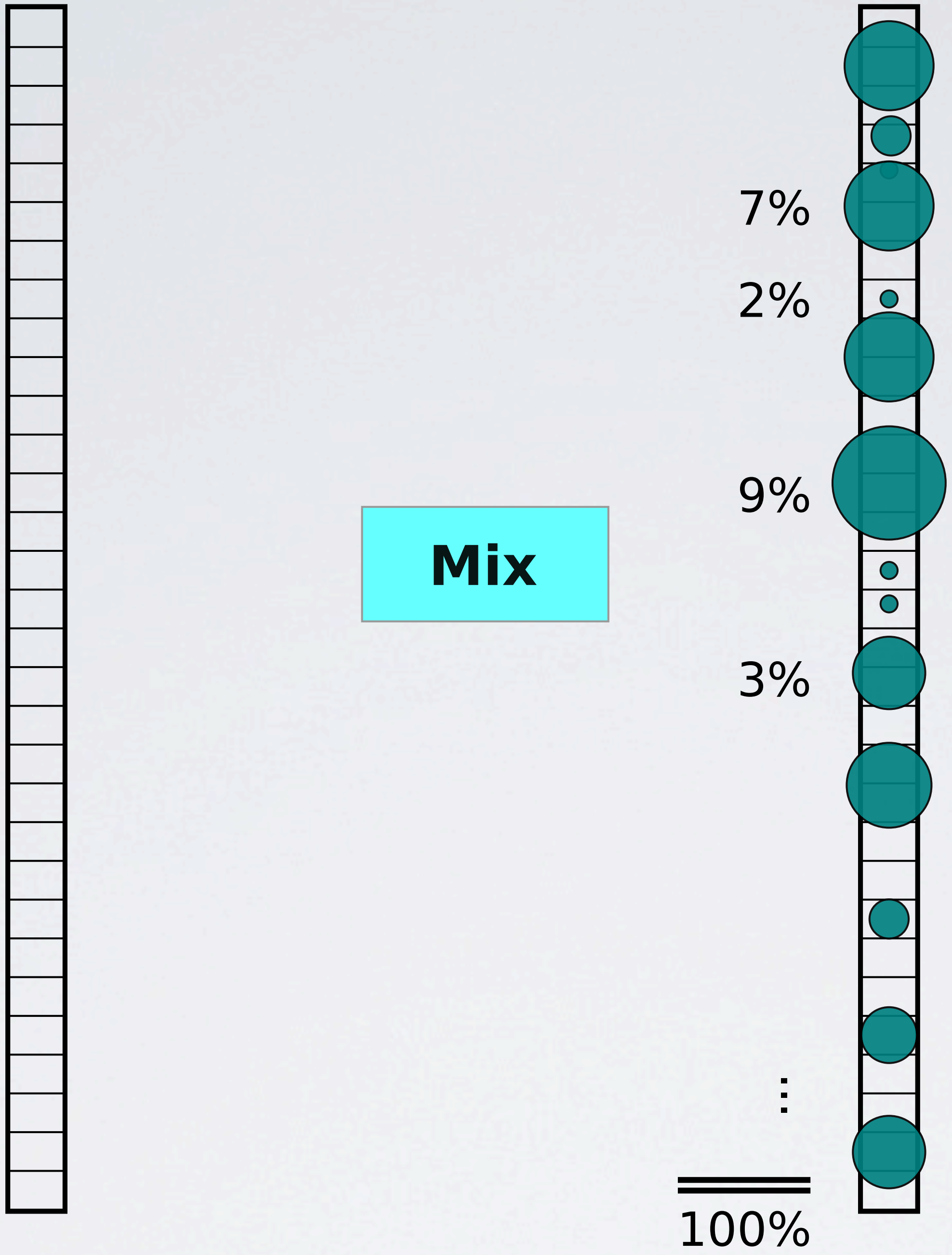
TIMING CORRELATION

Batching strategies reduce information leak

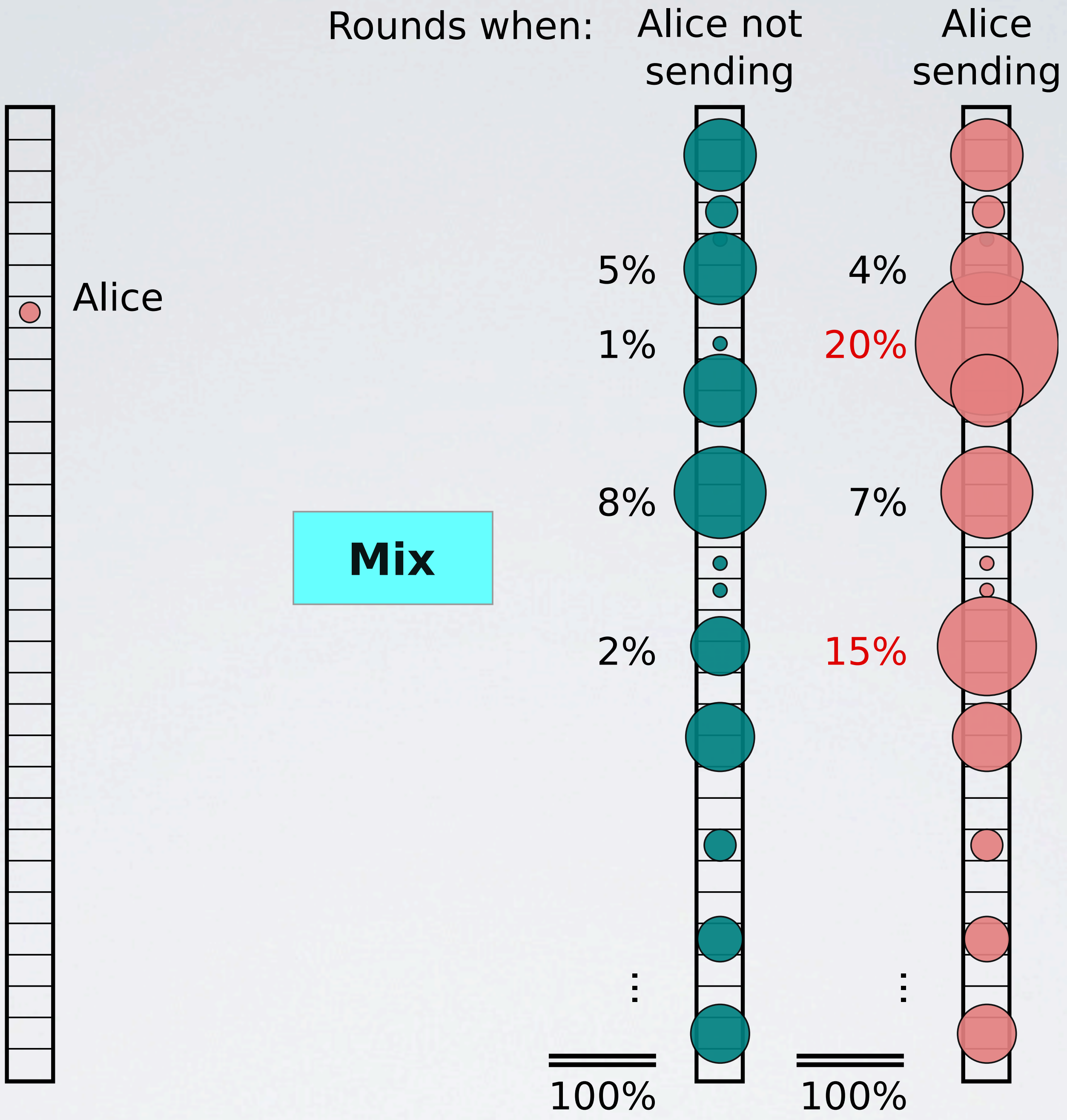


TIMING CORRELATION

Batching strategies reduce information leak



RECIPIENT PROFILING



TRAFFIC ANALYSIS

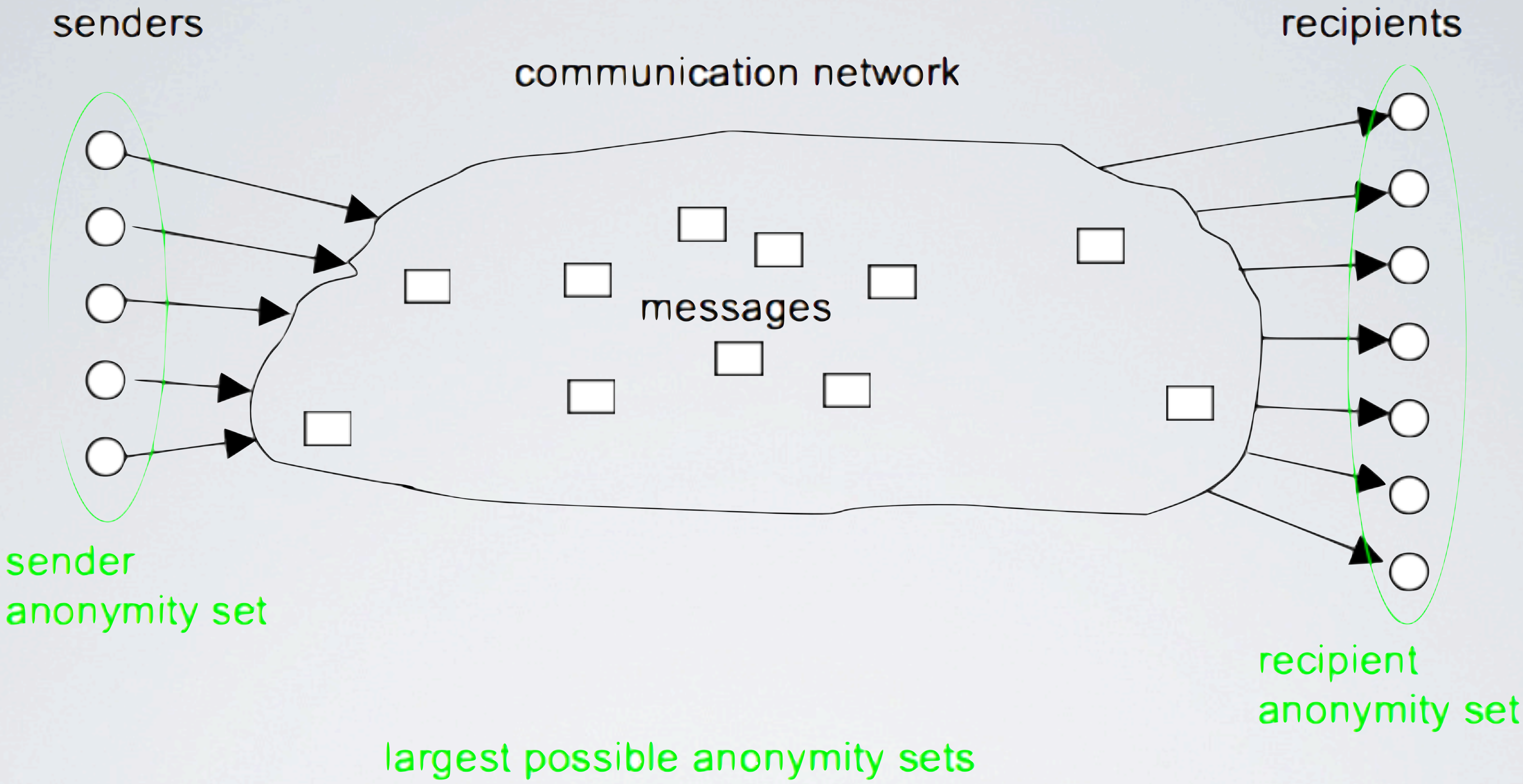


Fig. 3: Anonymity sets within the setting

A TERMINOLOGY FOR TALKING ABOUT PRIVACY BY DATA MINIMIZATION: ANONYMITY, UNLINKABILITY, UNDETECTABILITY, UNOBSERVABILITY,

Andreas Pfitzmann, Marit Hansen

(first published 2000, last updated 2010)

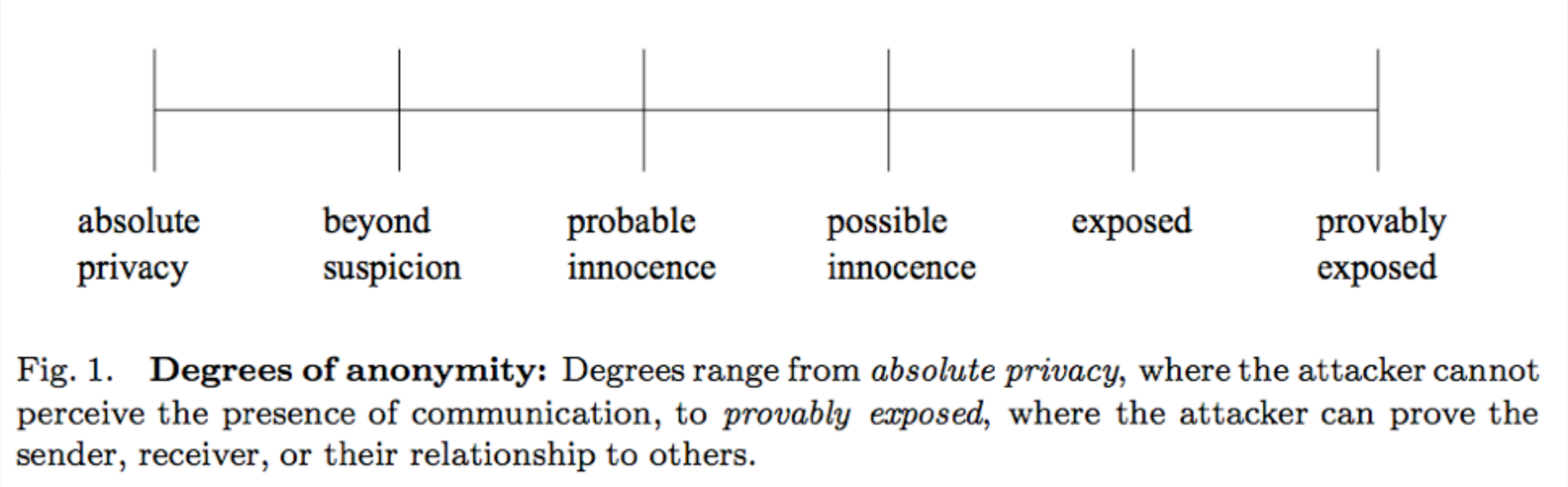
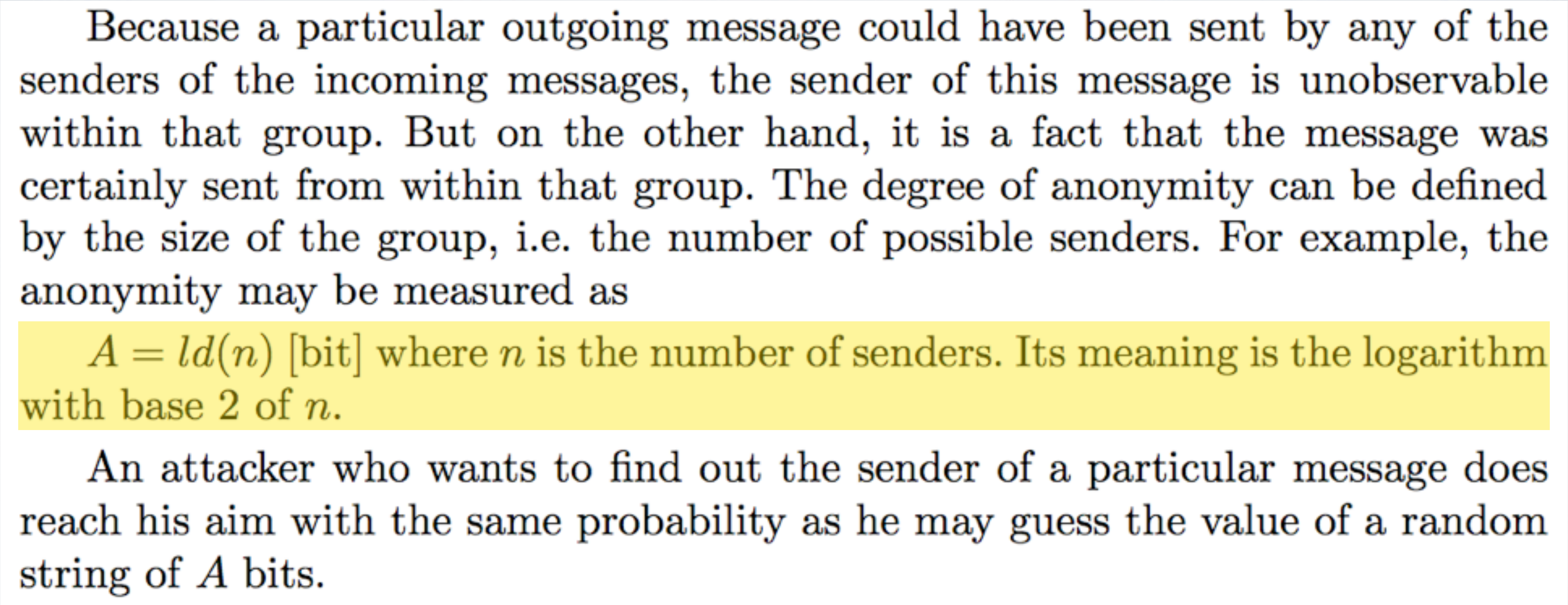


Fig. 1. **Degrees of anonymity:** Degrees range from *absolute privacy*, where the attacker cannot perceive the presence of communication, to *provably exposed*, where the attacker can prove the sender, receiver, or their relationship to others.

CROWDS: ANONYMITY FOR WEB TRANSACTIONS

Michael K. Reiter, Aviel D. Rubin (1997)



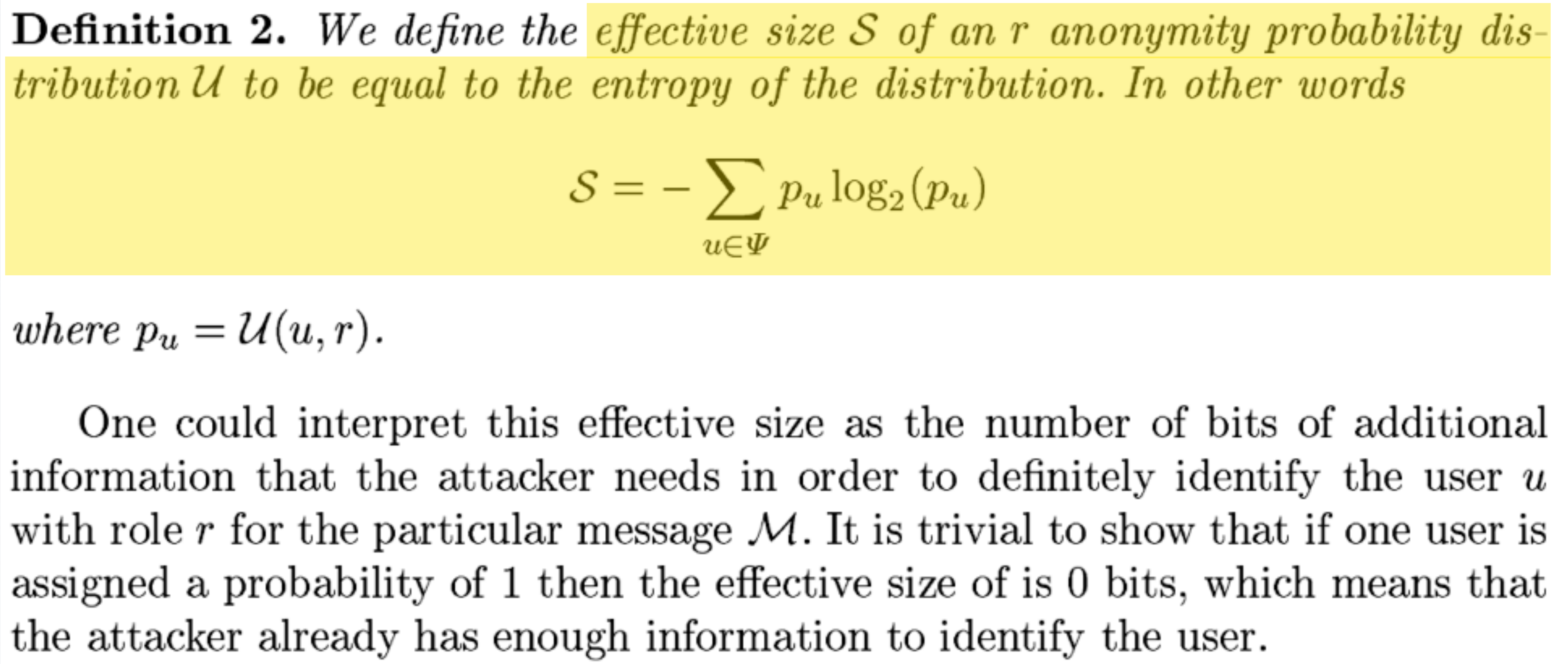
Because a particular outgoing message could have been sent by any of the senders of the incoming messages, the sender of this message is unobservable within that group. But on the other hand, it is a fact that the message was certainly sent from within that group. The degree of anonymity can be defined by the size of the group, i.e. the number of possible senders. For example, the anonymity may be measured as

$A = ld(n)$ [bit] where n is the number of senders. Its meaning is the logarithm with base 2 of n.

An attacker who wants to find out the sender of a particular message does reach his aim with the same probability as he may guess the value of a random string of A bits.

THE DISADVANTAGES OF FREE MIX ROUTES AND HOW TO OVERCOME THEM

Oliver Berthold, Andreas Pfitzmann, Ronny Standtke (2001)



Definition 2. *We define the effective size $\mathcal{S}$ of an r anonymity probability distribution $\mathcal{U}$ to be equal to the entropy of the distribution. In other words*

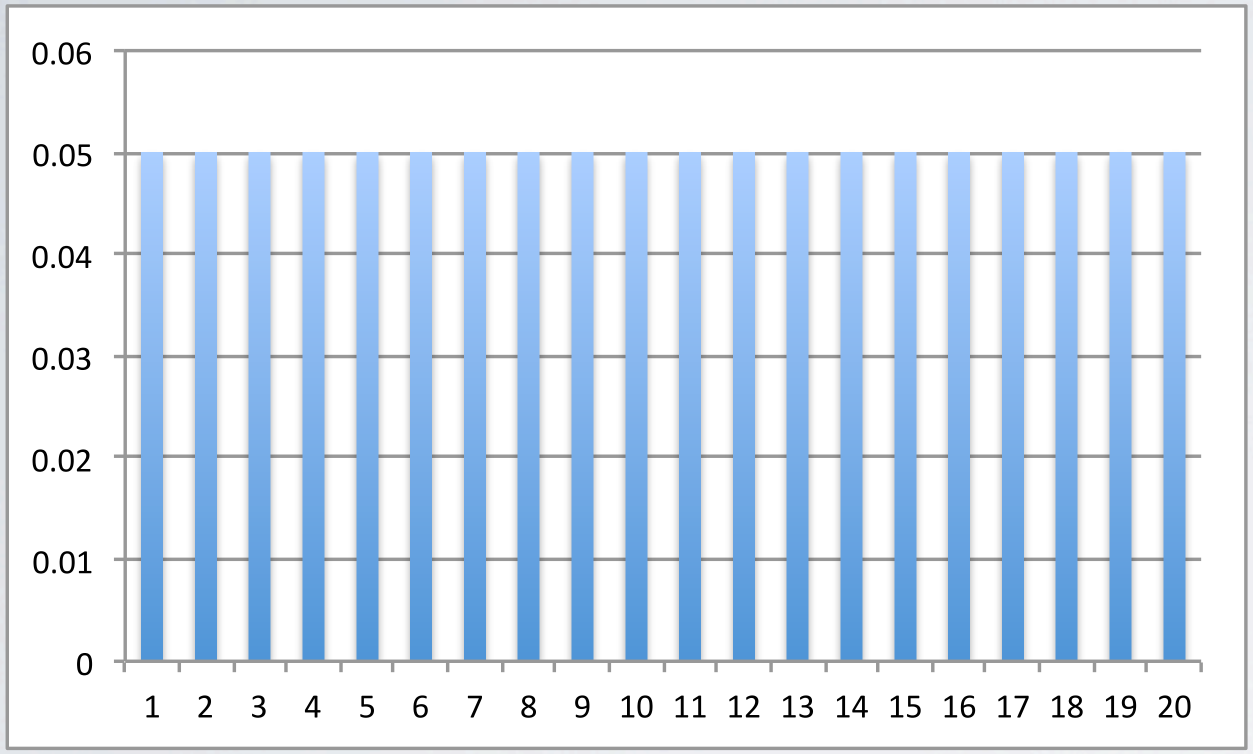
$$\mathcal{S} = -\sum_{u \in \Psi} p_u \log_2(p_u)$$

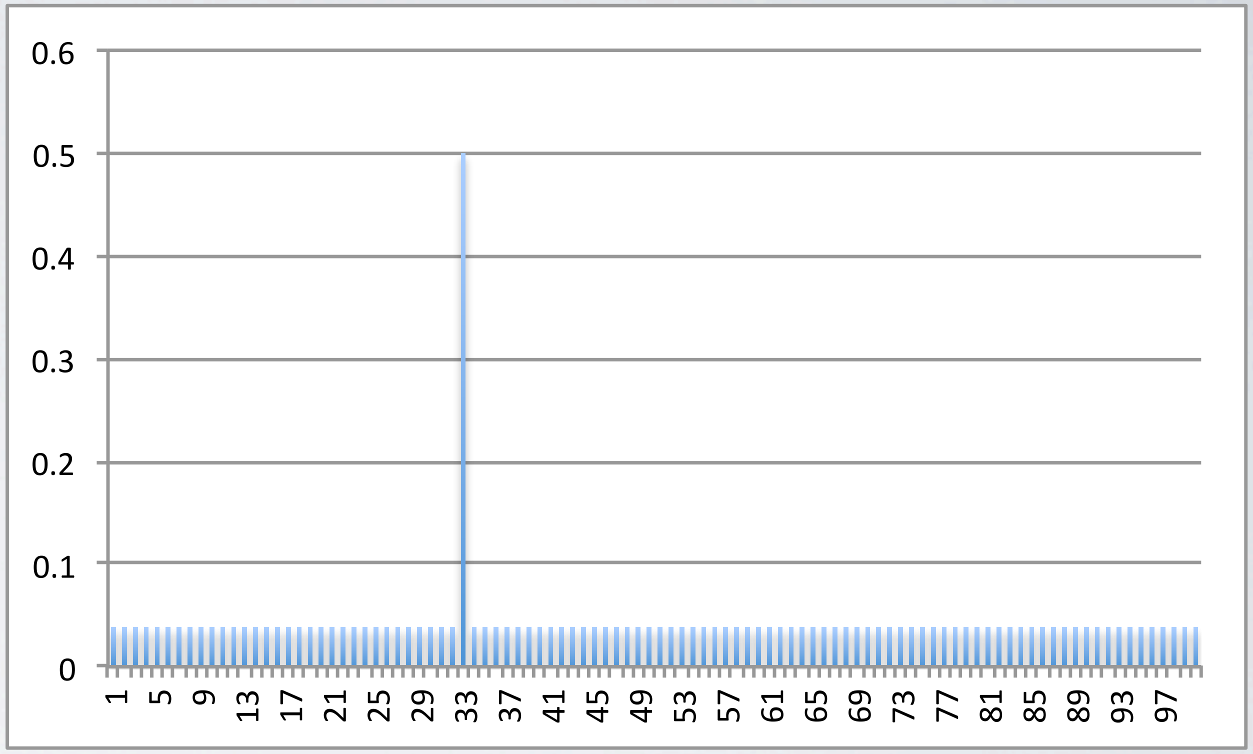
where $p_u = \mathcal{U}(u, r)$.

One could interpret this effective size as the number of bits of additional information that the attacker needs in order to definitely identify the user u with role r for the particular message $\mathcal{M}$. It is trivial to show that if one user is assigned a probability of 1 then the effective size of is 0 bits, which means that the attacker already has enough information to identify the user.

TOWARDS AN INFORMATION THEORETIC METRIC FOR ANONYMITY

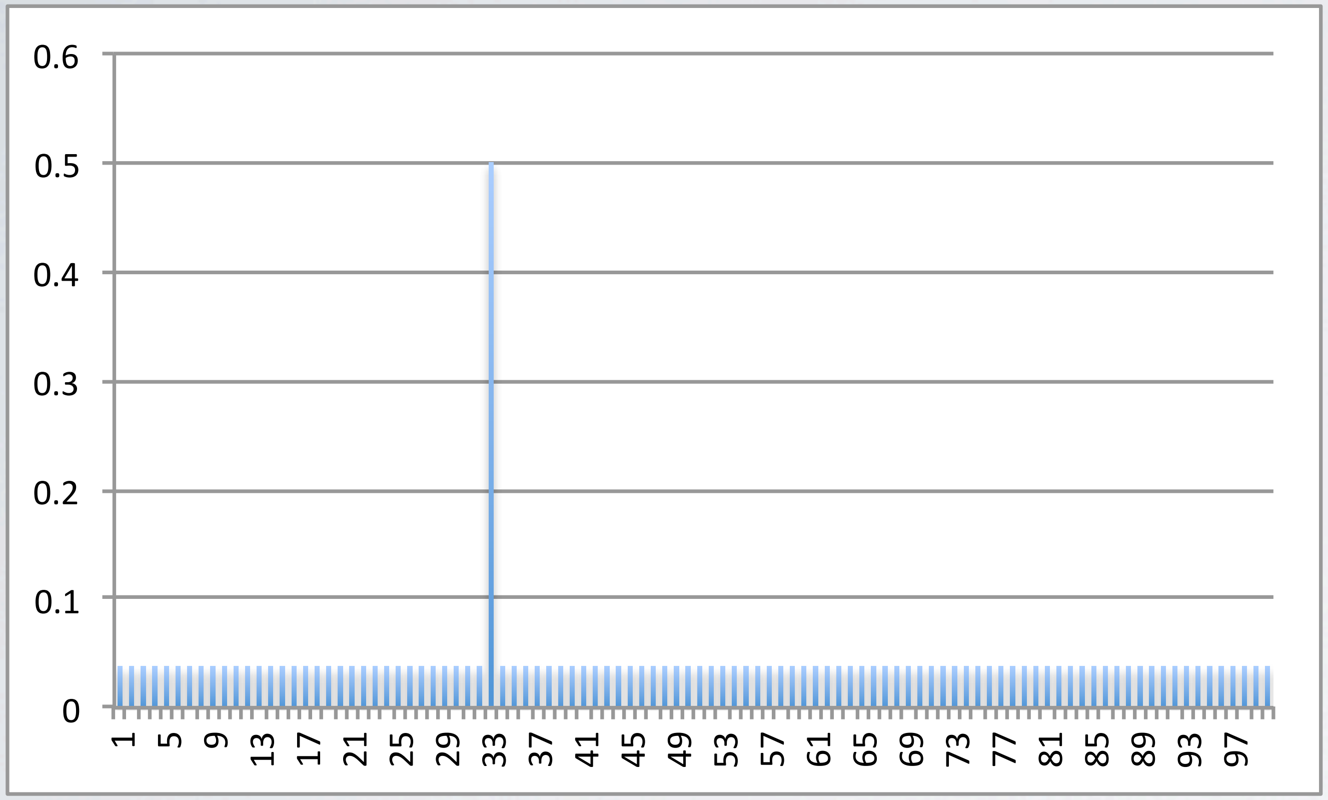
Andrei Serjantov, George Danezis (2002)


0.06
0.05
0.04
0.03
0.02
0.01
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

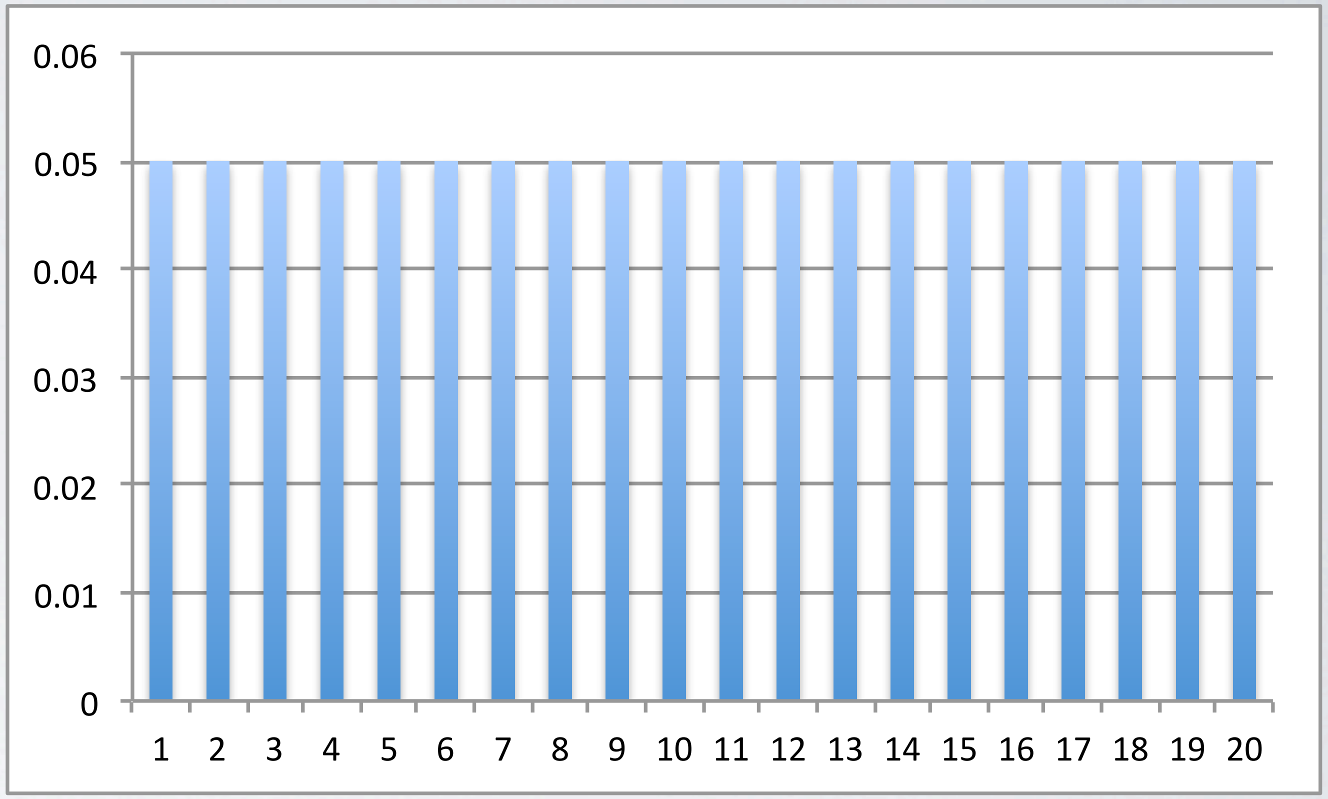

0.6
0.5
0.4
0.3
0.2
0.1
0
1 5 9 13 17 21 25 29 33 37 41 45 49 53 57 61 65 69 73 77 81 85 89 93 97

Entropy = 4.3

Normalized entropy = 0.6

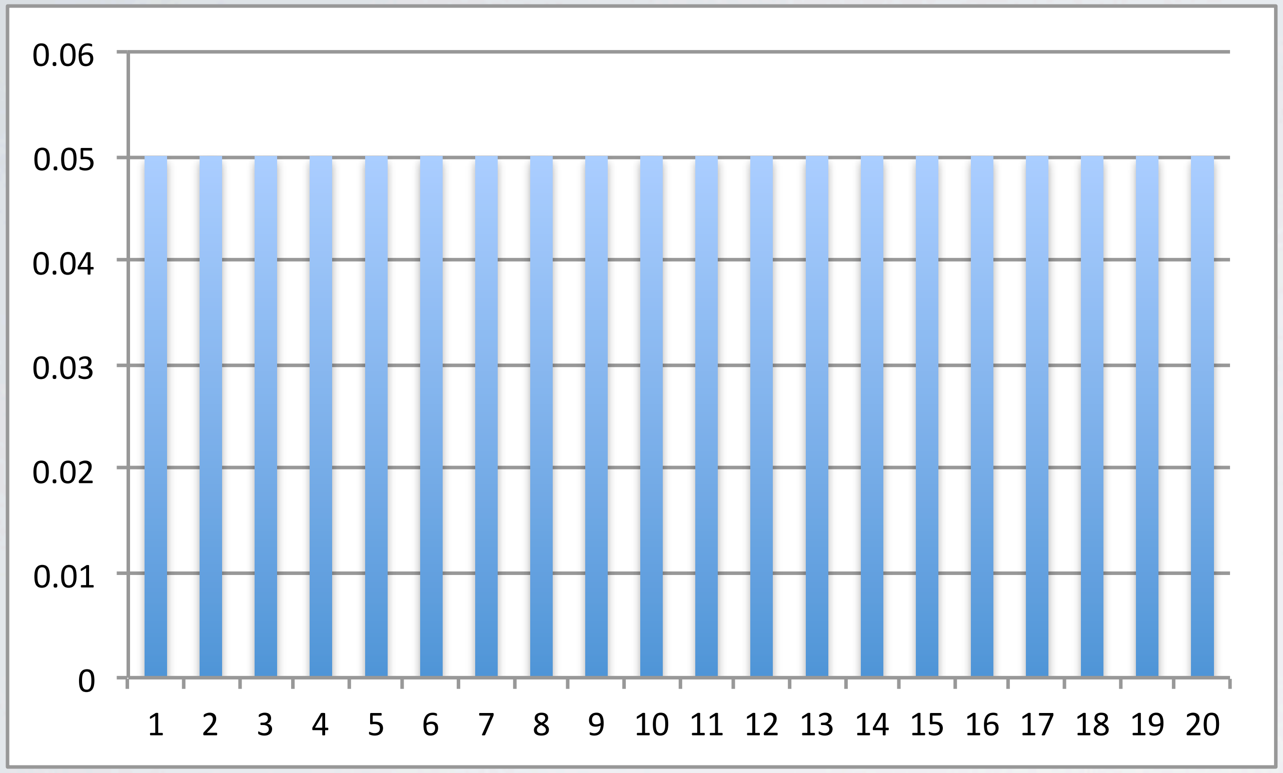


Normalized entropy = 1

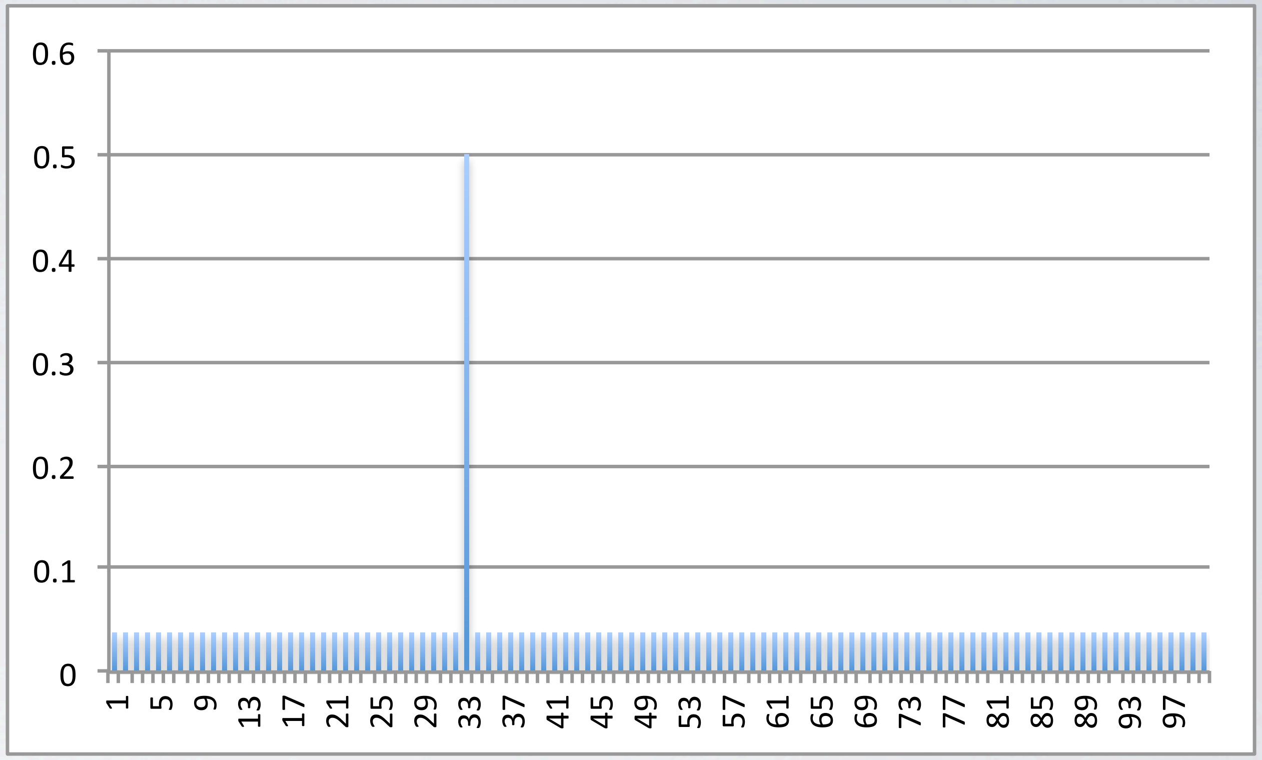


Degree of anonymity

Beyond suspicion



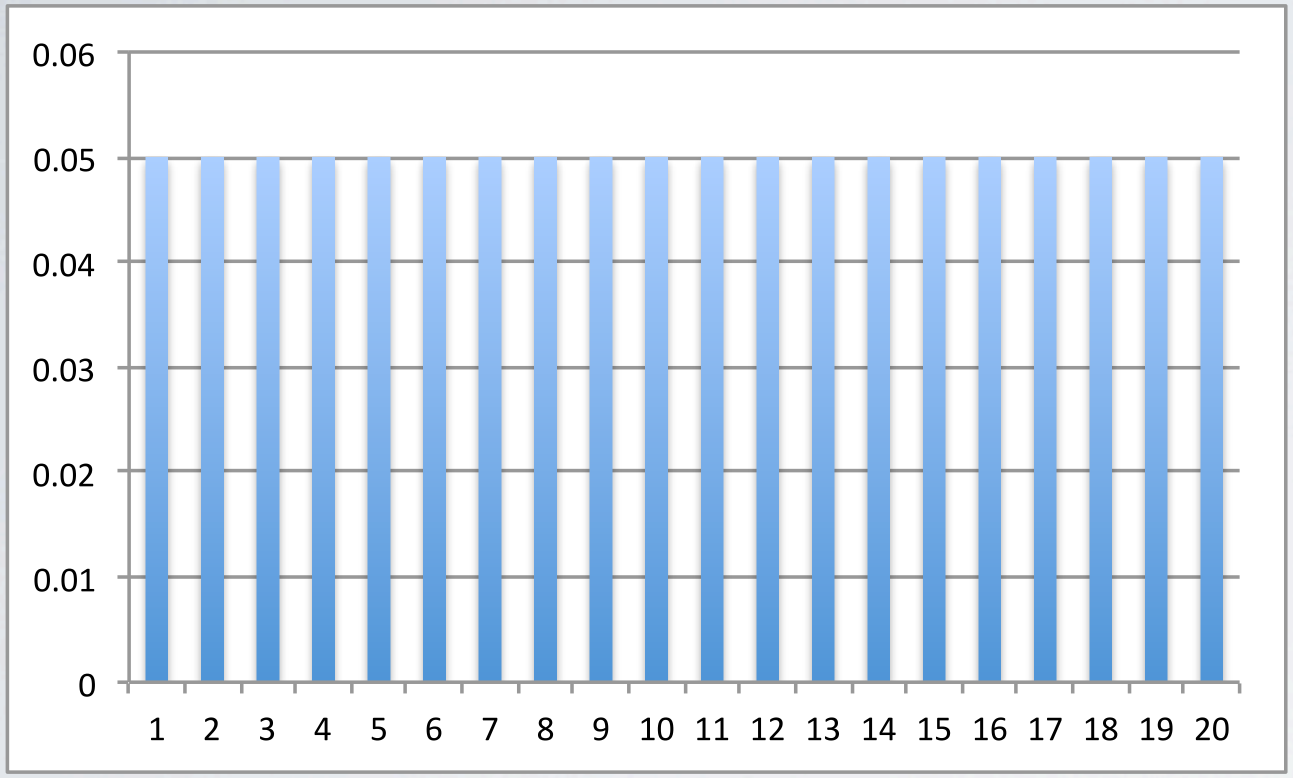
Probable innocence



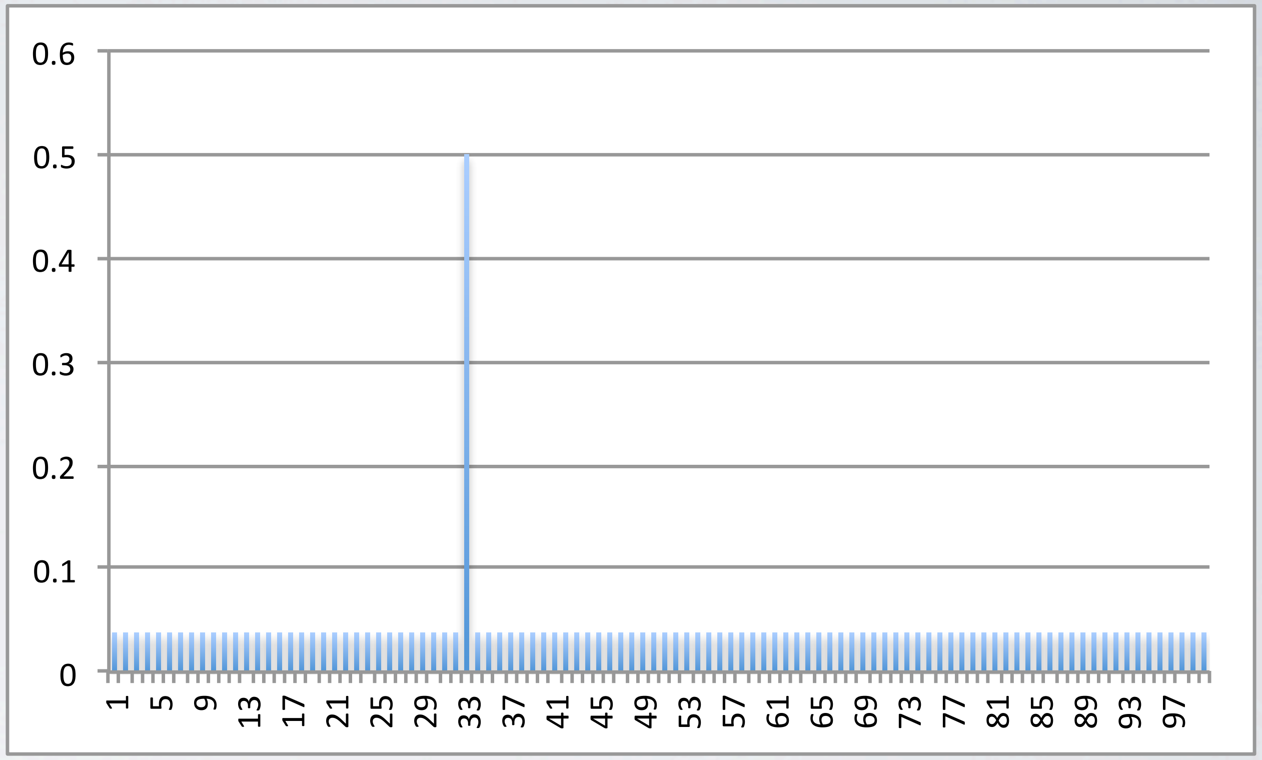
WHICH ONE IS BETTER?

Anonymity set size

20



101

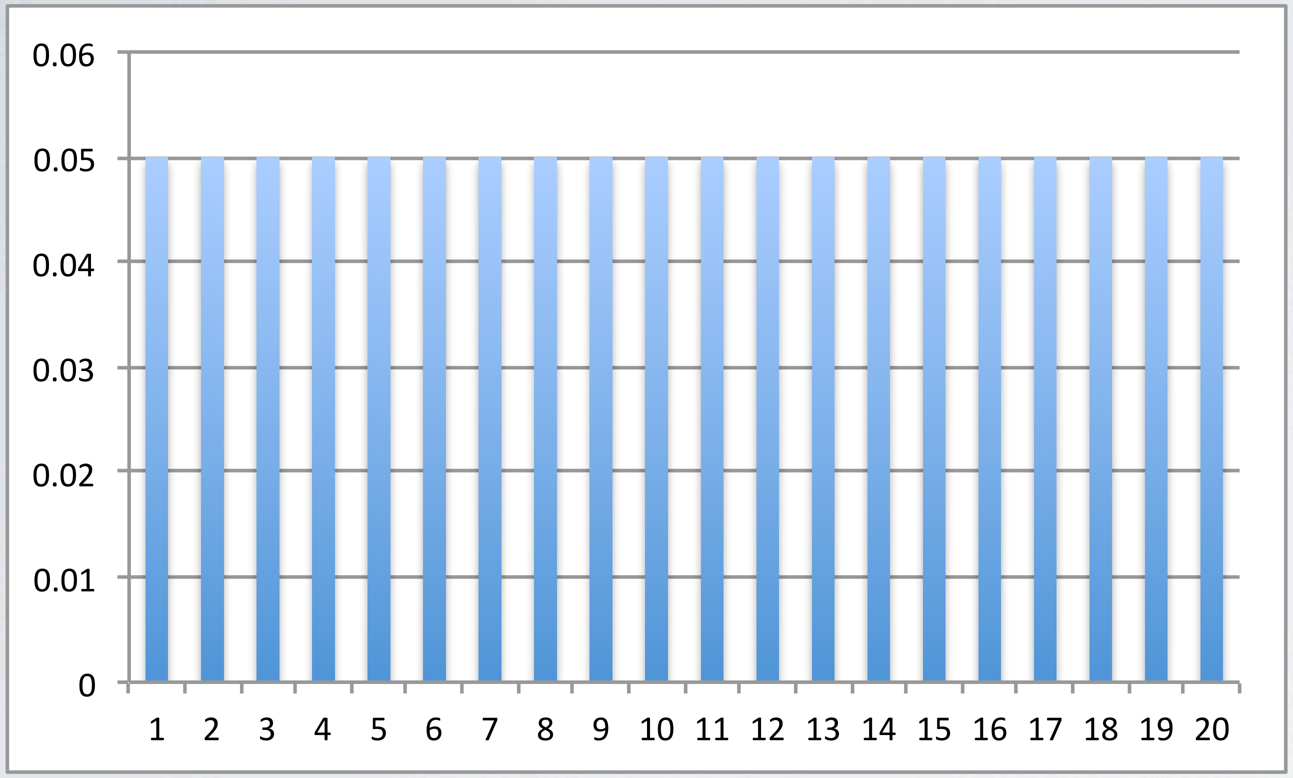


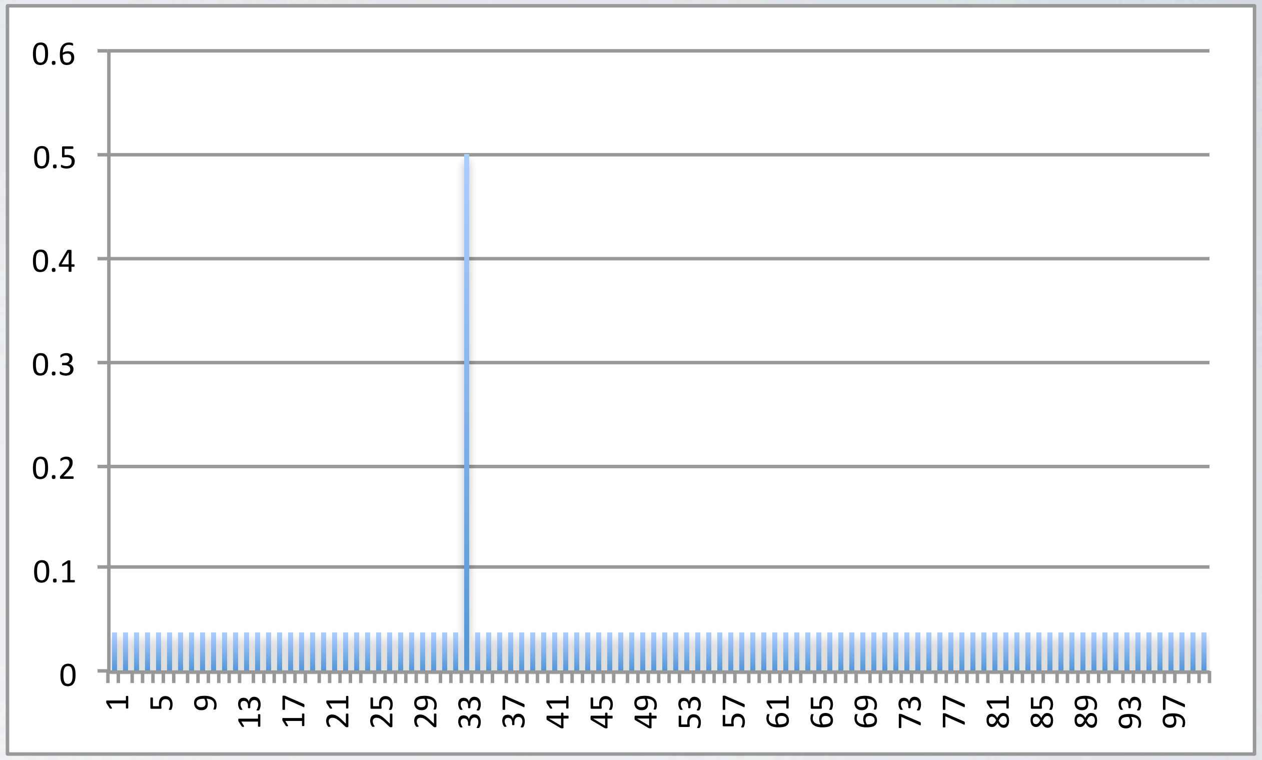
WHICH ONE IS BETTER?

Entropy

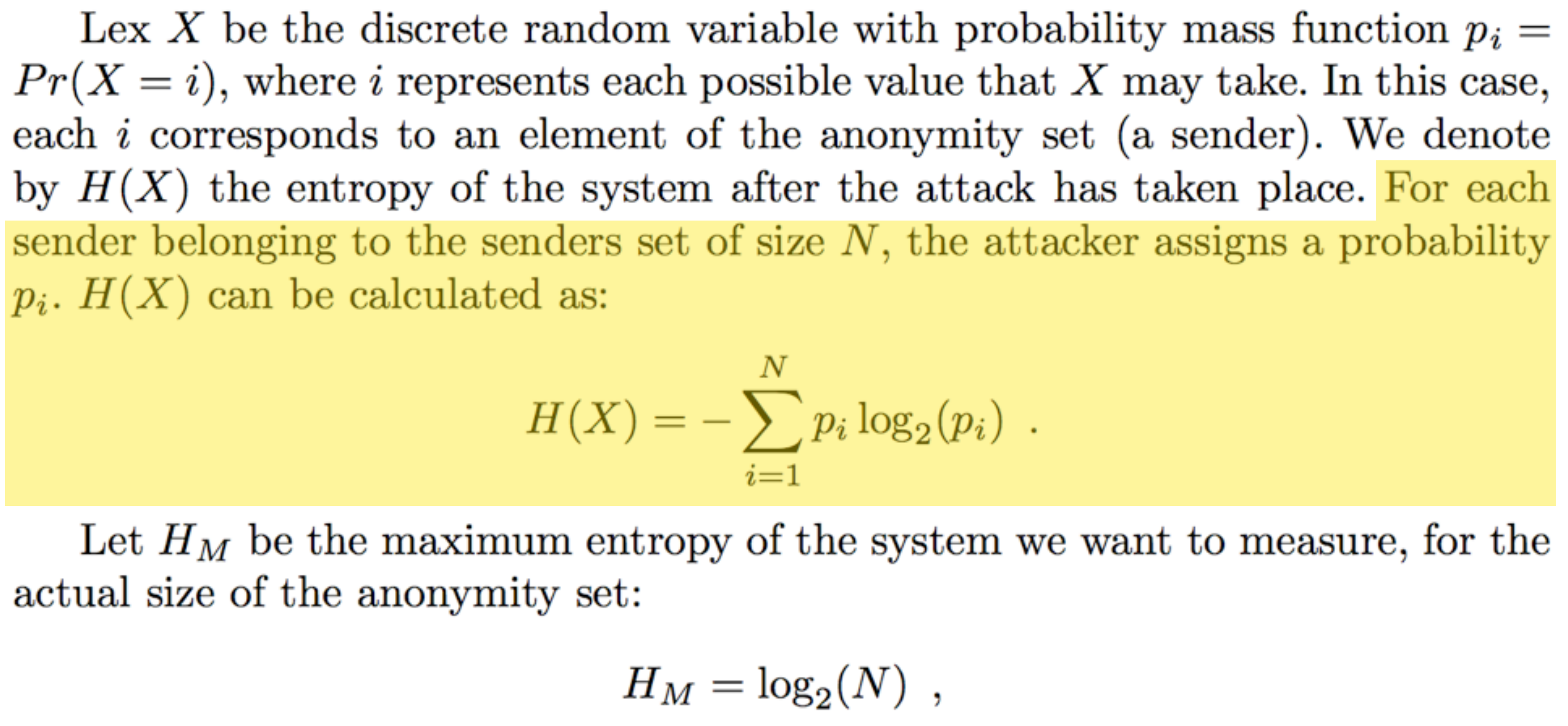
4.3

4.3





WHICH ONE IS BETTER?



Lex X be the discrete random variable with probability mass function $p_i = Pr(X = i)$, where i represents each possible value that X may take. In this case, each i corresponds to an element of the anonymity set (a sender). We denote by $H(X)$ the entropy of the system after the attack has taken place. For each sender belonging to the senders set of size N, the attacker assigns a probability p_i. $H(X)$ can be calculated as:

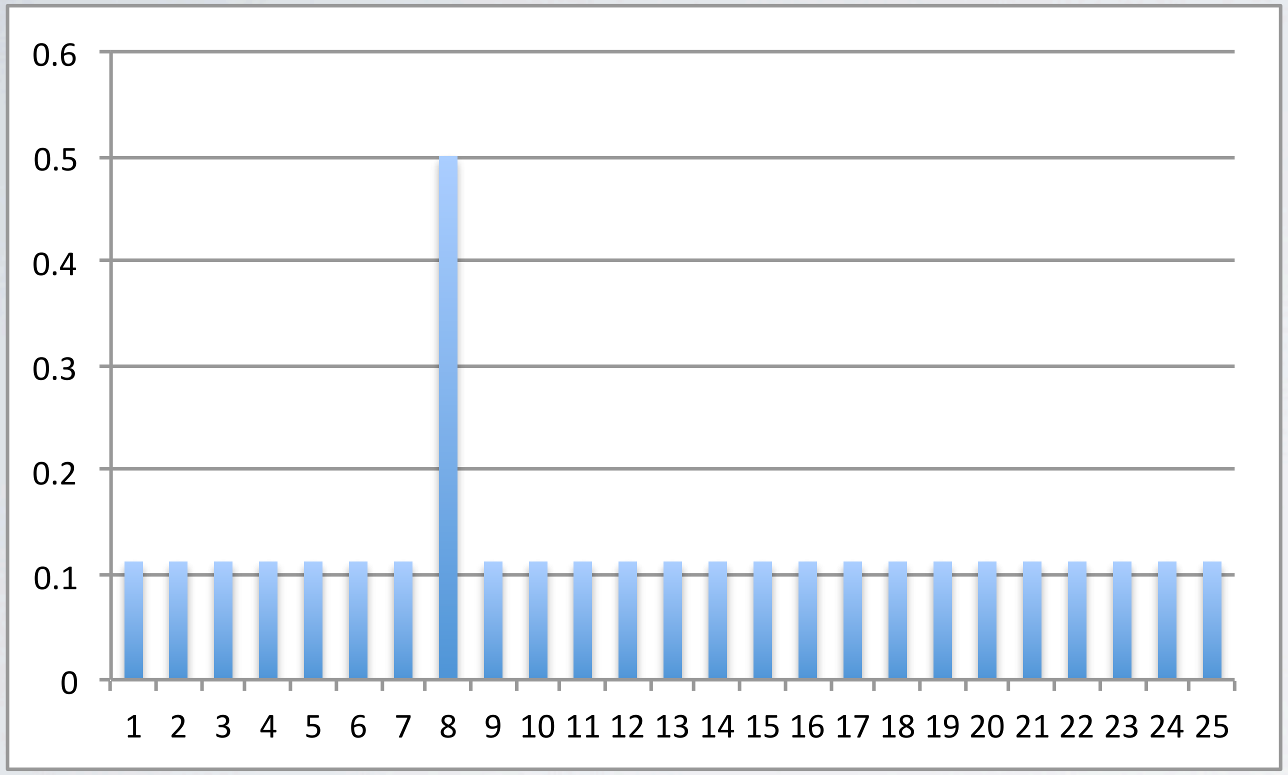
$$H(X) = -\sum_{i=1}^{N} p_i \log_2(p_i) \ .$$

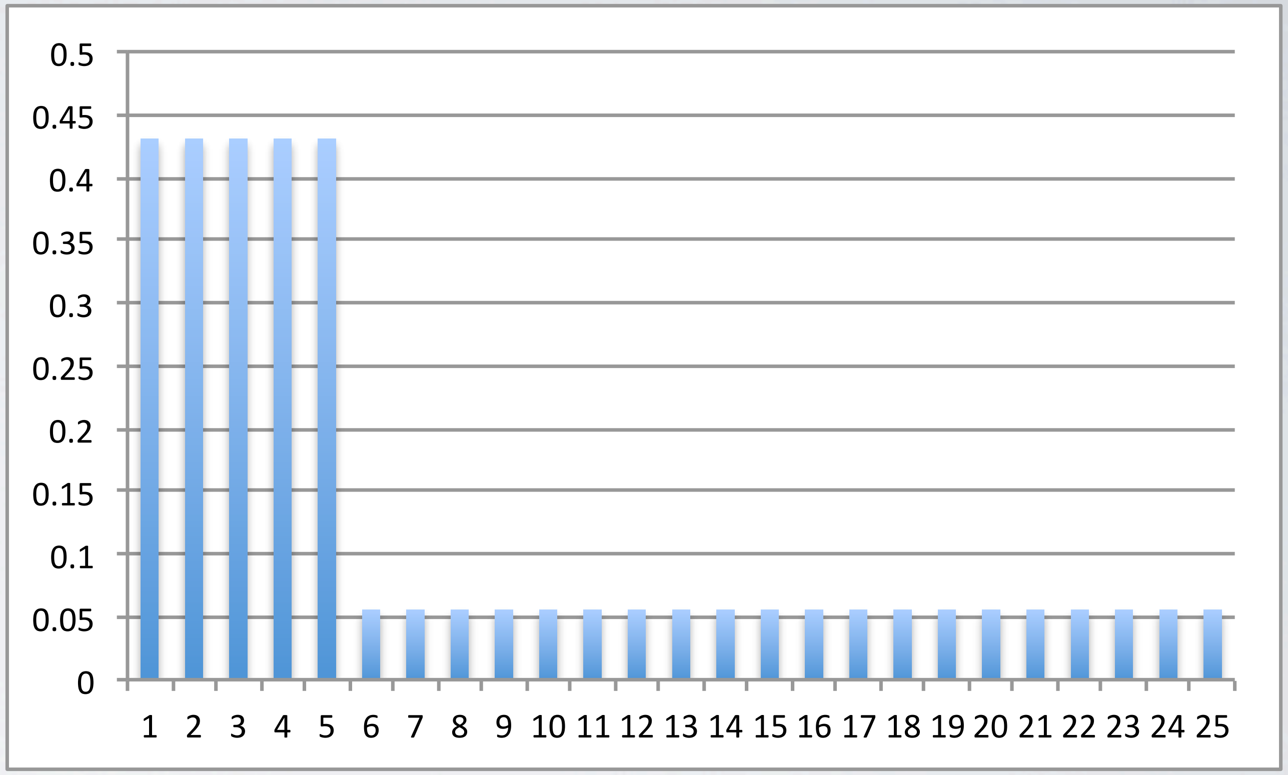
Let H_M be the maximum entropy of the system we want to measure, for the actual size of the anonymity set:

$$H_M = \log_2(N) \ ,$$

TOWARDS MEASURING ANONYMITY

Claudia Diaz, Stefaan Seys, Joris Claessens, Bart Preneel (2002)



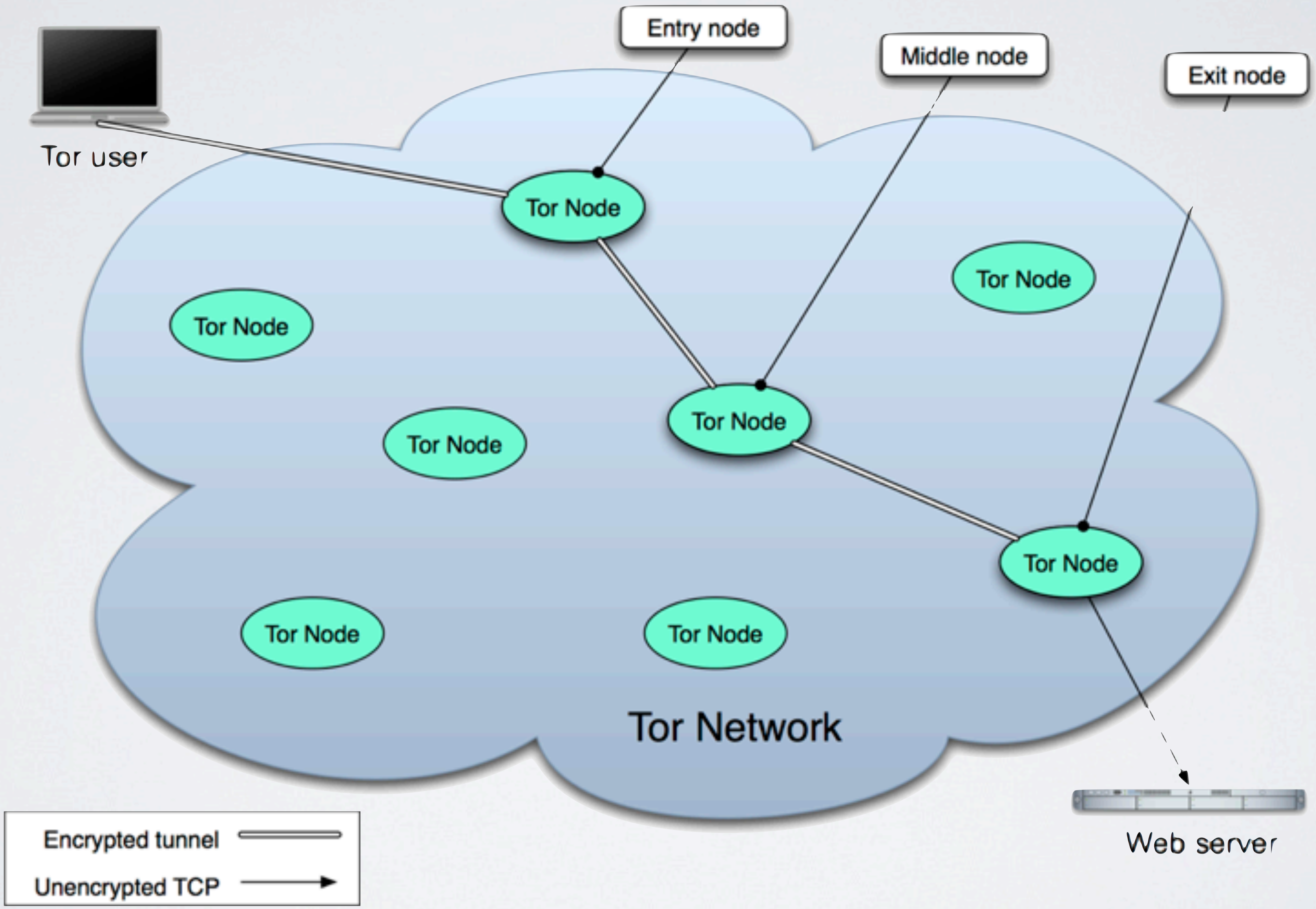


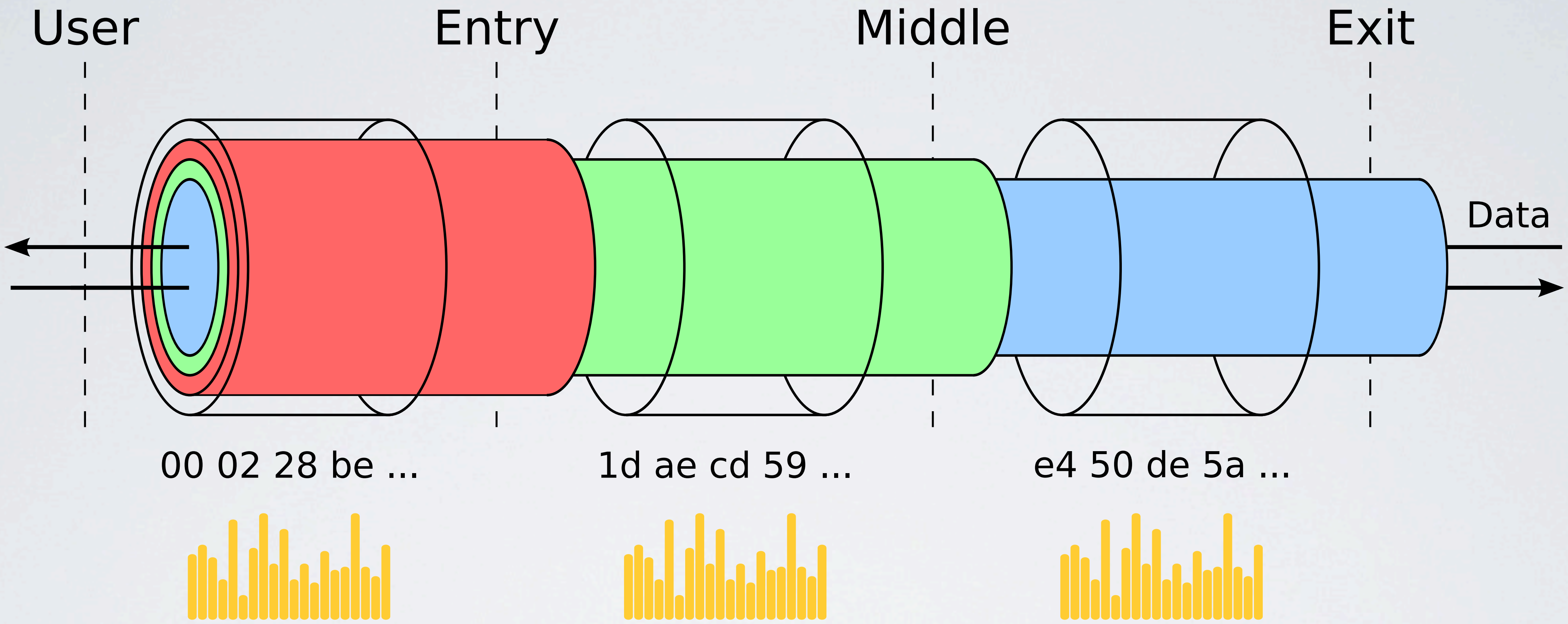
Entropy = 3.2
Normalized entropy = 0.7

ENTROPY-BASED ANONYMITY METRICS

- Four metrics
- Four answers on what is best
- Generalization possible (worst-case entropy, Rényi entropy)
- Need to think about threat model
  - What is the budget
  - What is the goal

TOR


Entry node
Middle node
Exit node
Tor user
Tor Node
Tor Node
Tor Node
Tor Node
Tor Node
Tor Node
Tor Node
Tor Node
Tor Network
Web server
Encrypted tunnel
Unencrypted TCP



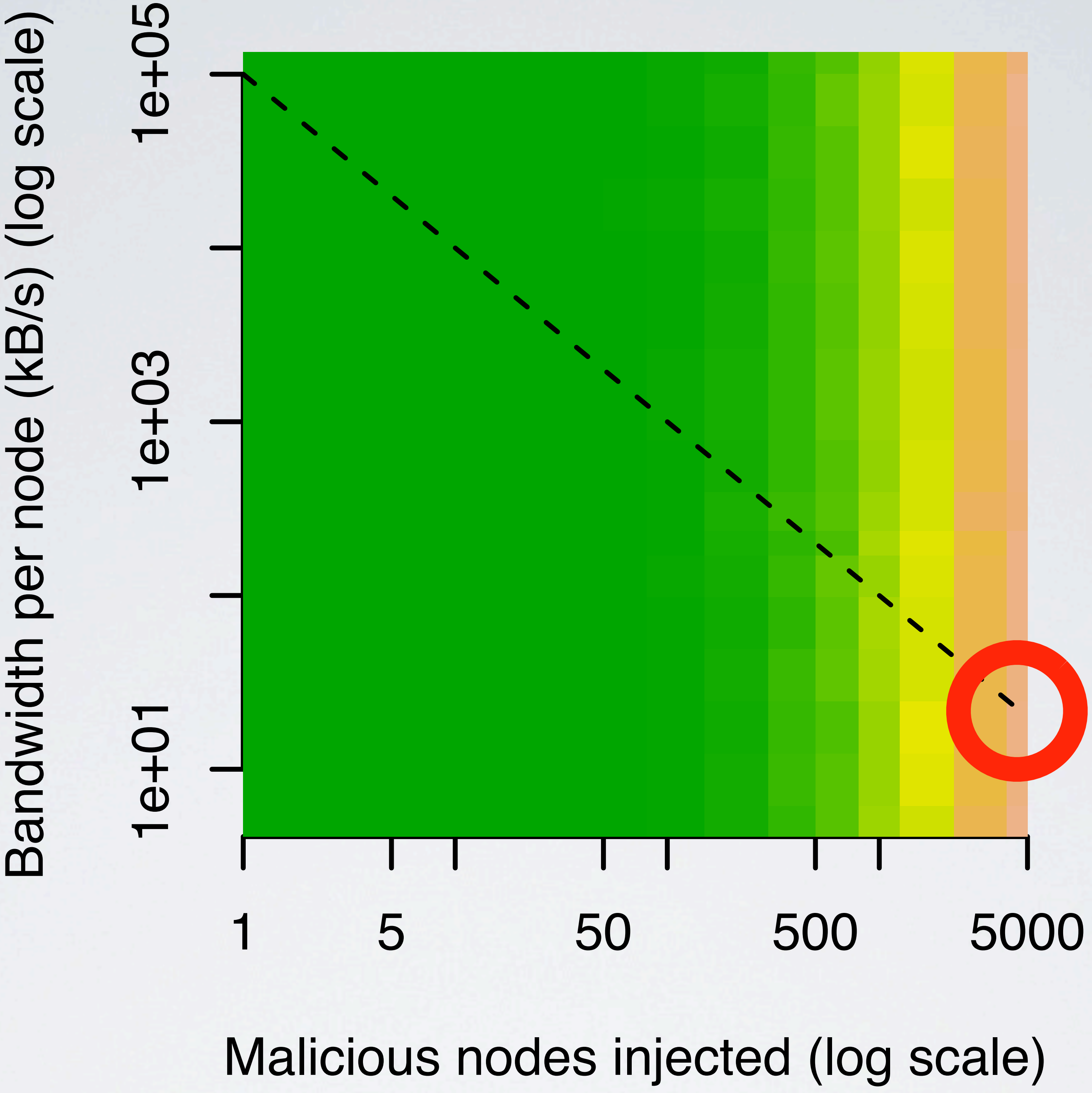
TOR KEY EXCHANGE

TOR THREAT MODEL

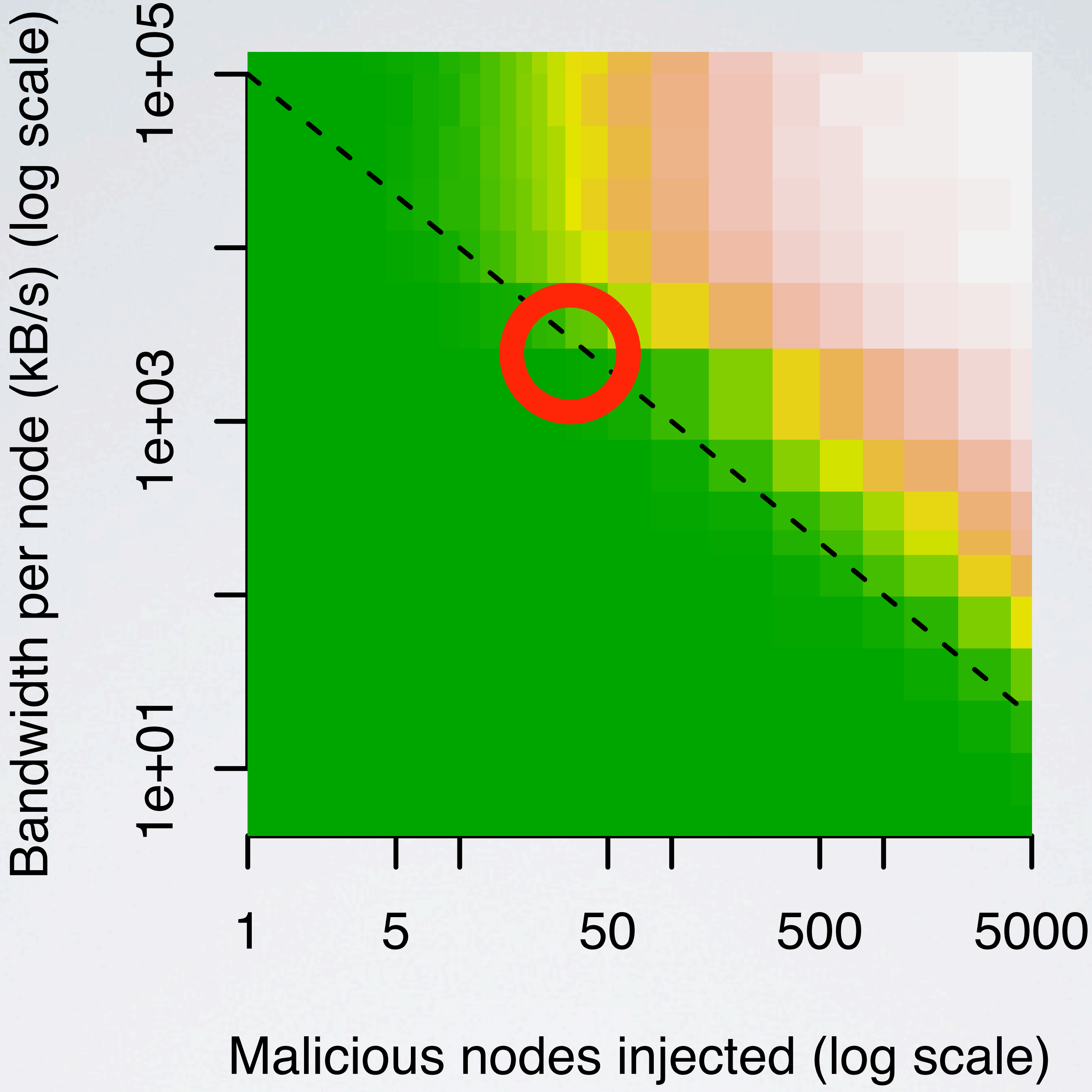
- Traditional assumption is global-passive
- Both too strong and too weak
- Few adversaries are global
- Weak adversaries can be active
- Application of entropy to node selection not best approach

DIRECT ANALYSIS APPROACH

- Model attacker space
- Model attack for all possible attack
- Network security level is best available attacker strategy



UNIFORM NODE SELECTION



BANDWIDTH WEIGHTED

CENSORSHIP RESISTANCE

- Increasing number of Tor users want resistance to censorship
- How do we evaluate proposed approaches
- Need to look at costs of adversaries
  - CPU, memory, losing face, ...
- Need to look at benefit to Tor
  - More users, more countries, consistent performance

CONCLUSIONS

- Metrics need to be developed hand in hand with threat models

- If metric doesn’t allow threat model to be a parameter then which one is implied

- Evolution of anonymity metrics can illustrate some approaches and techniques of calculating them may generalize