
Anonymity and Privacy in Electronic Services

Claudia Díaz

Katholieke Universiteit Leuven
Dept. Electrical Engineering – ESAT/COSIC

Promoters: Prof. Preneel
Prof. Vandewalle

December 19, 2005, Leuven-Heverlee

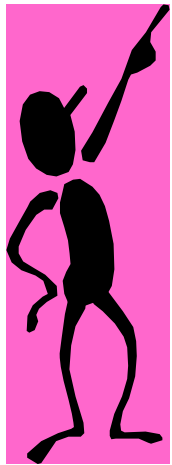
Outline

- **Introduction to Anonymity**
- Anonymity Metrics
- Mixes
- Passive Attacks - Outline
- Active Attacks
- Practical Evaluations - Outline
- Contributions and Open Questions

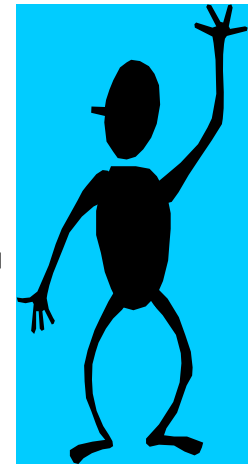
Applications of Anonymity (Towards Recipient or Third Parties)

- E-Voting (anonymity + receipt-freeness)
- E-Health, Help Lines, Social Services
- Protection against Profiling
 - Unfair Commercial Practices
- User-Controlled Identity Management
 - Prevention of Identity Theft
- Whistle Blowing
- Freedom of Speech
- Privacy-Enhancing Technologies

Classical Security Model



Alice



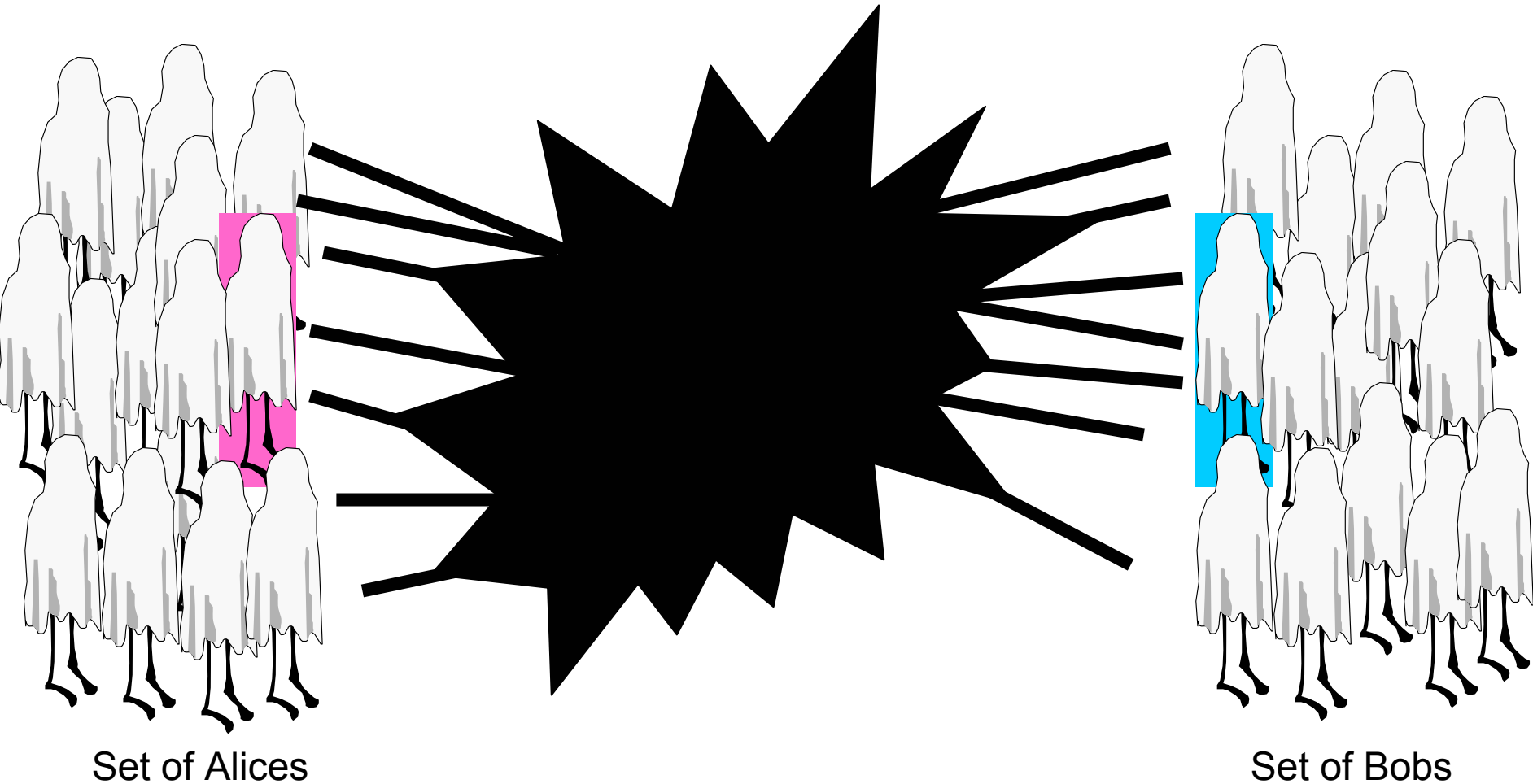
Bob



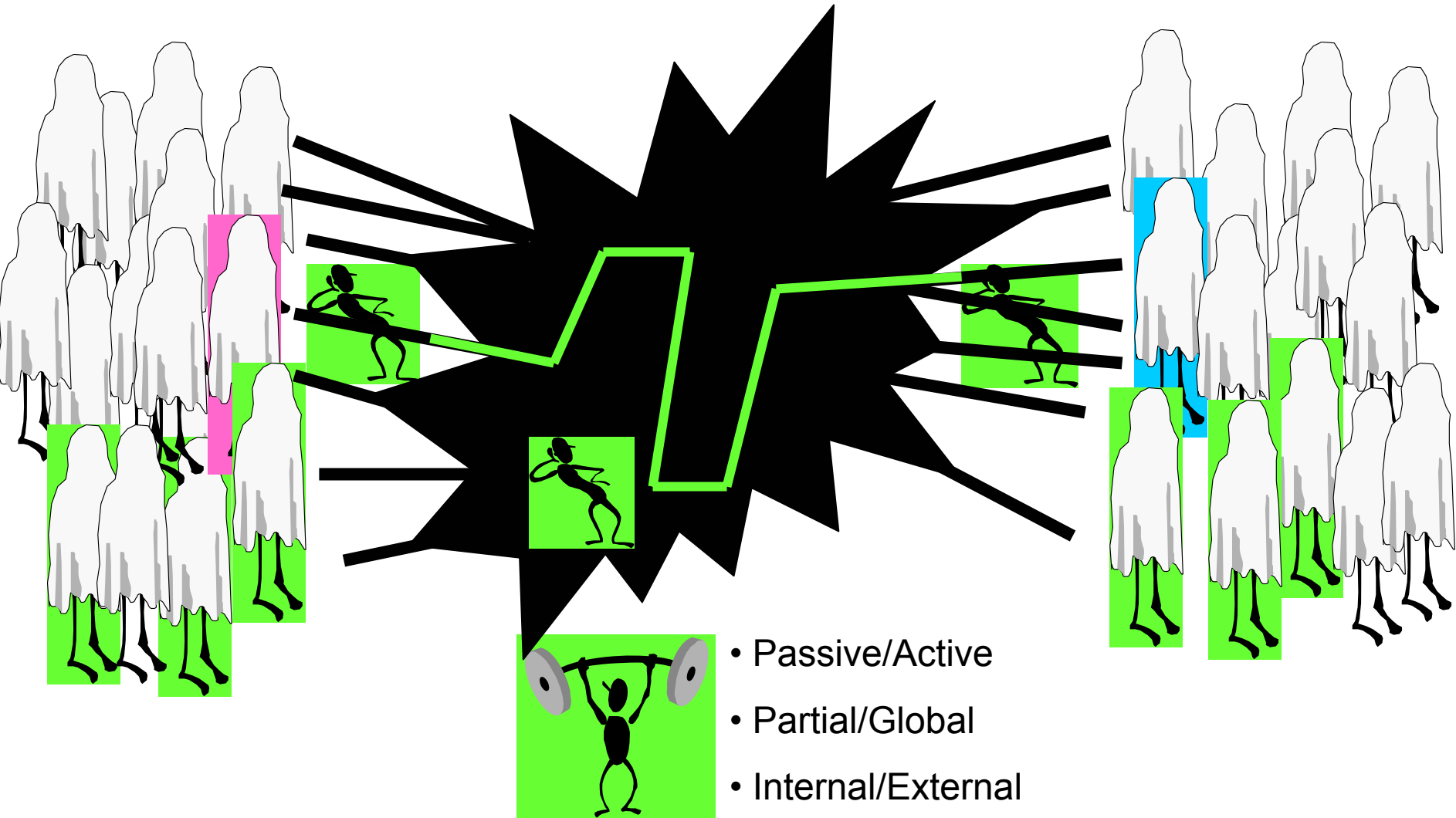
Eve

Passive / Active

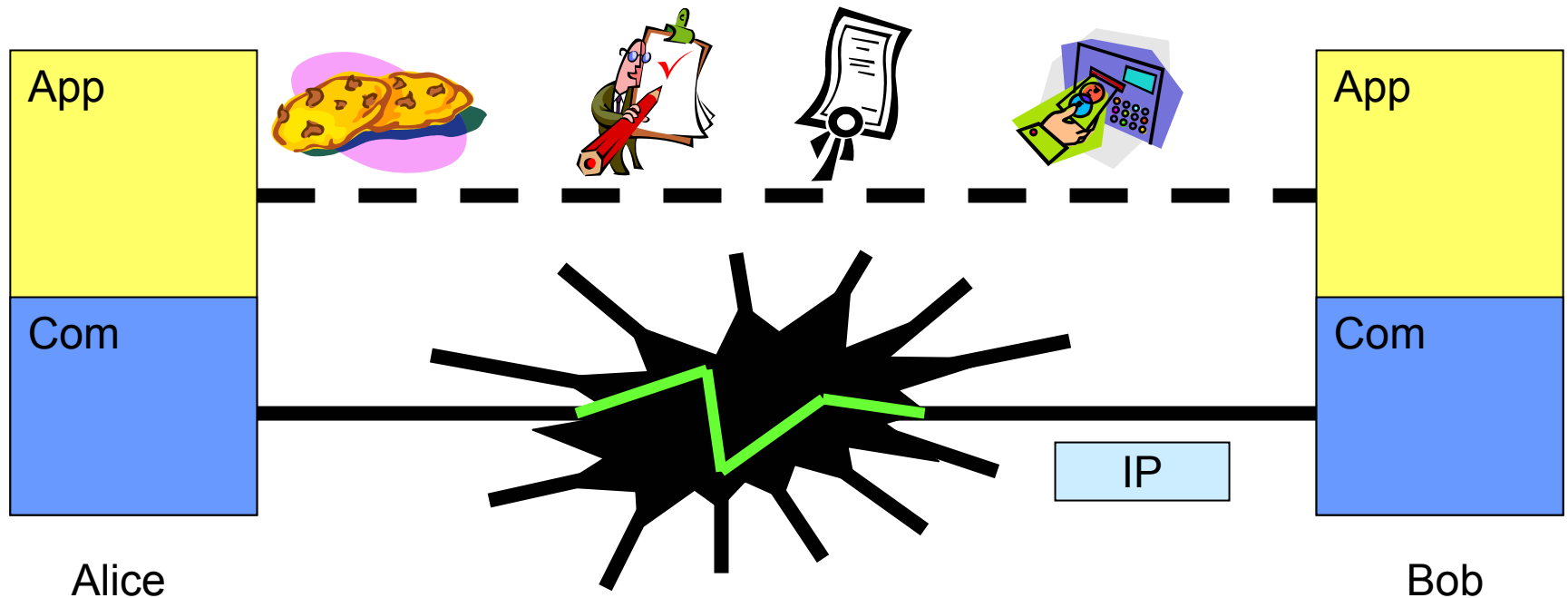
Anonymity – Concept and Model



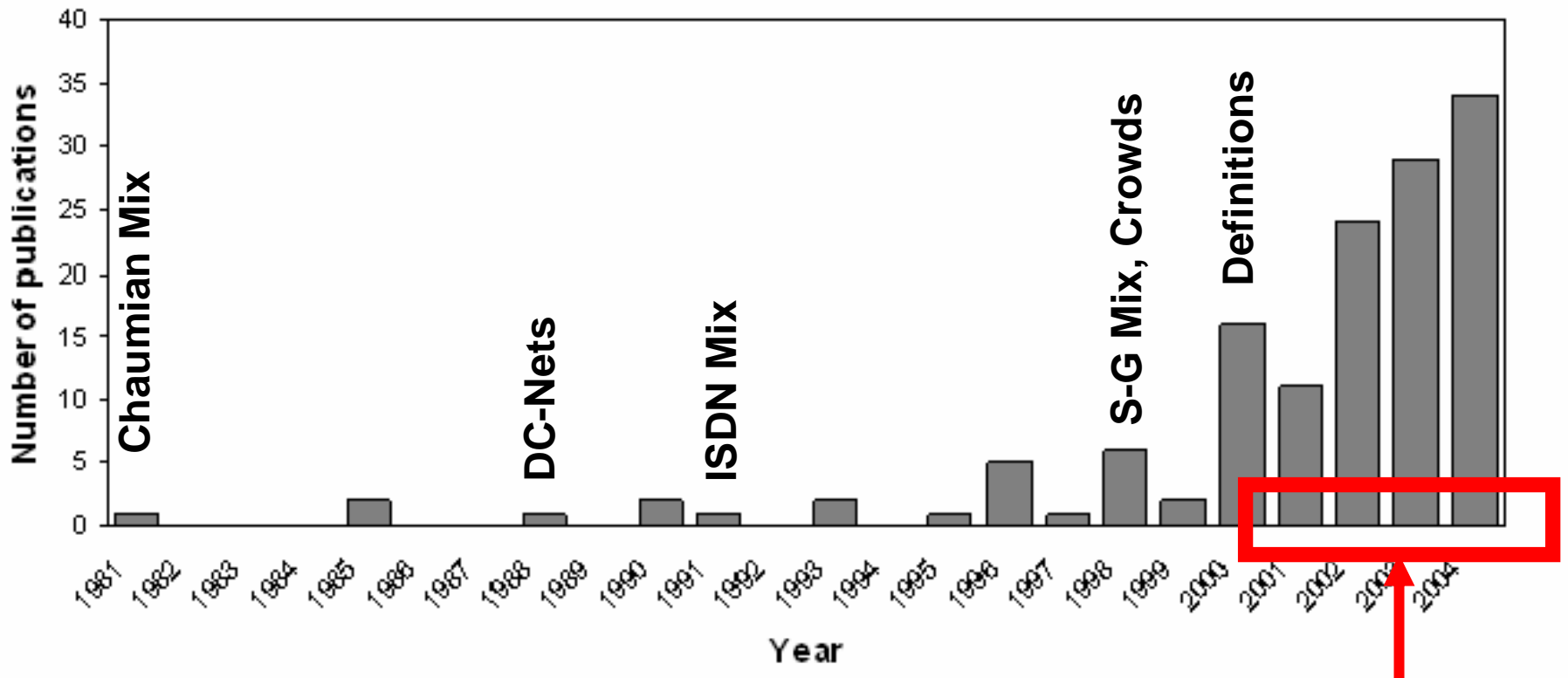
Anonymity Adversary



Anonymity – Data and Communication Layers



Related Work - Timeline



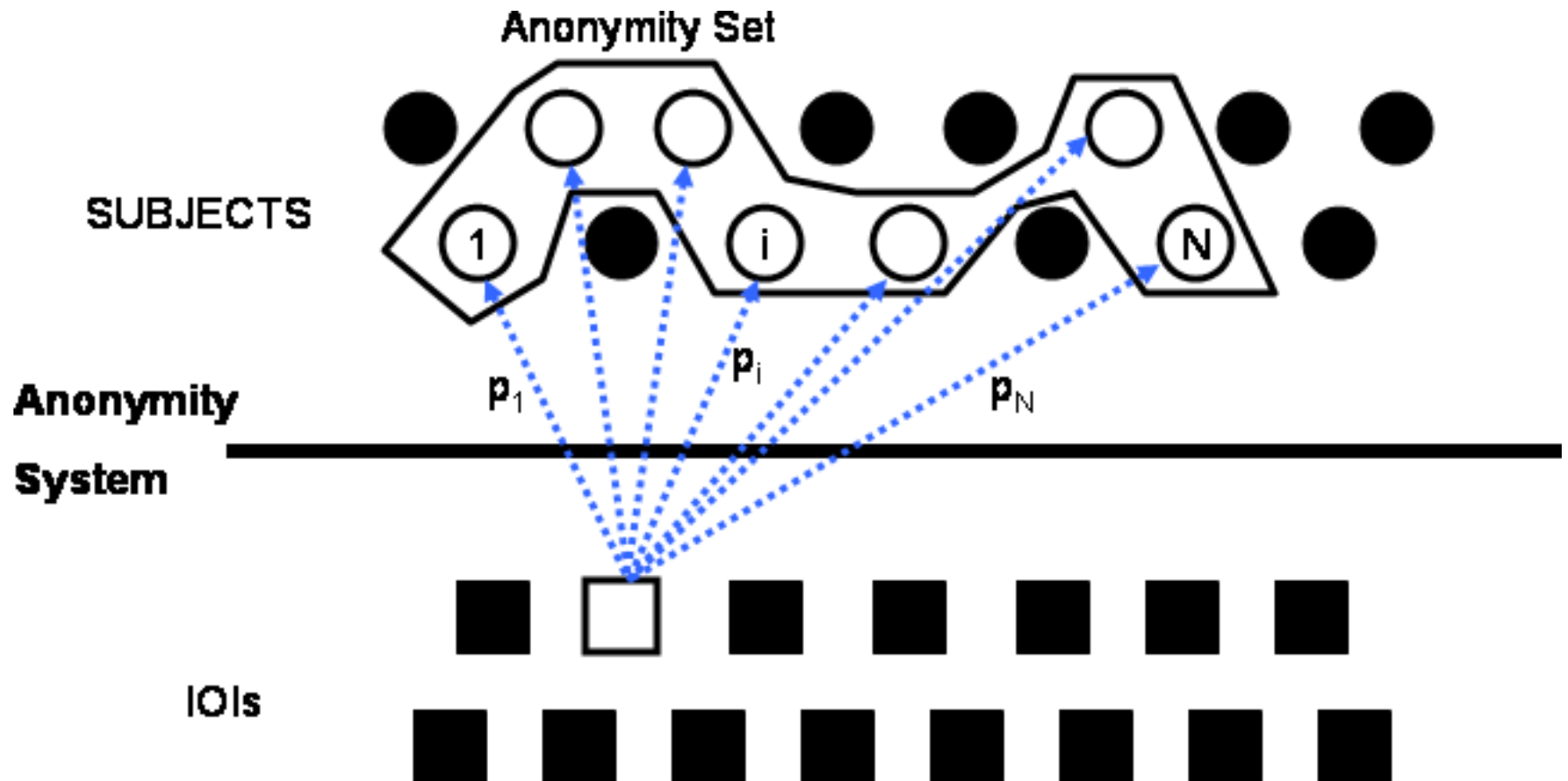
Source: Freehaven Anonymity Bibliography

Our Research

Outline

- Introduction to Anonymity
- **Anonymity Metrics**
- Mixes
- Passive Attacks - Outline
- Active Attacks
- Practical Evaluations - Outline
- Contributions and Open Questions

Abstract Model for Anonymity



Definition [PfiHan2000]

- ***First clear definition of anonymity (2000)***
 - ***Anonymity is the state of being not identifiable within a set of subjects, the anonymity set.***
 - The ***anonymity set*** is the set of all possible subjects who might cause an action or be addressed.

Entropy

- Measure of the amount of *information* required on the average to describe the random variable
- Measure of the *uncertainty* of a random variable
- Increases with N and with uniformity of distribution

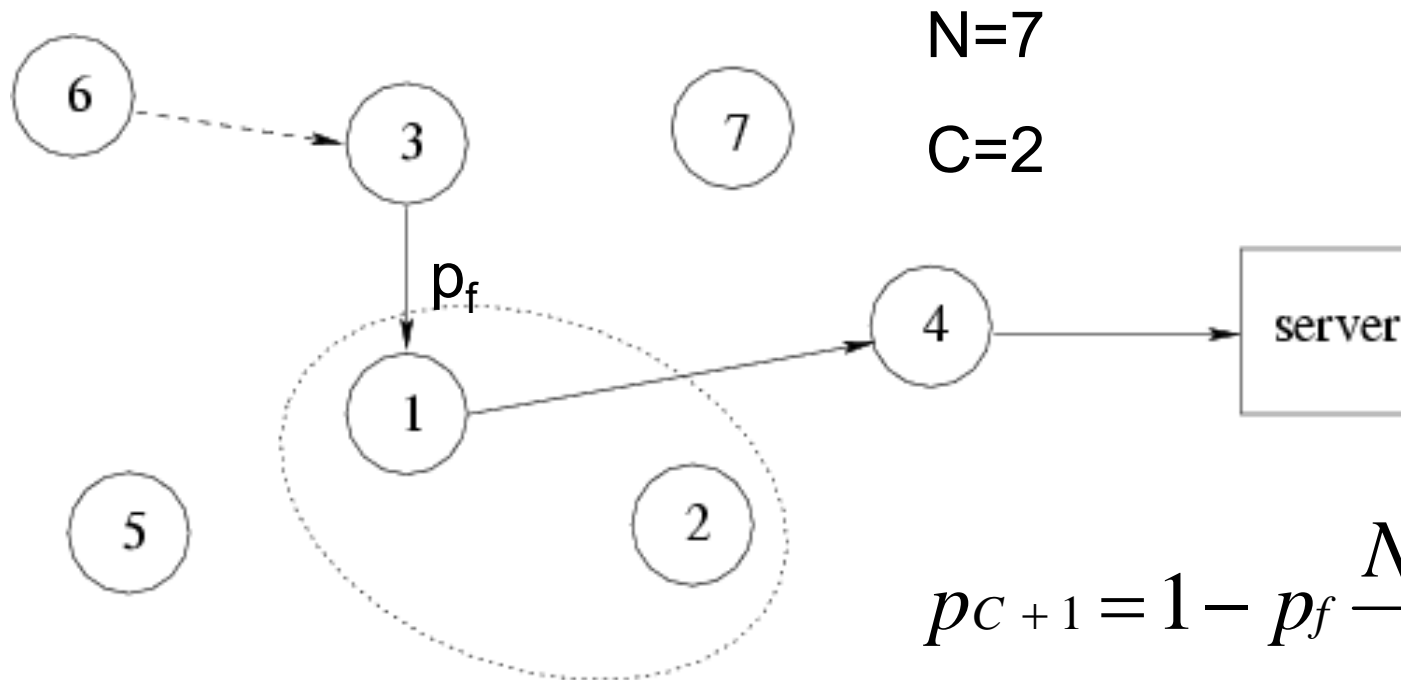
$$H = -\sum_{i=1}^N p_i \cdot \log_2(p_i)$$

Degree of anonymity

- How Much Anonymity Does the System Provide?
- “Independent” of the number of users
- Compare to the best the system can do: $H_M = \log_2(N)$
- Normalized Entropy

$$d = \frac{H}{H_M} \quad 0 \leq d \leq 1$$

Example: Crowds



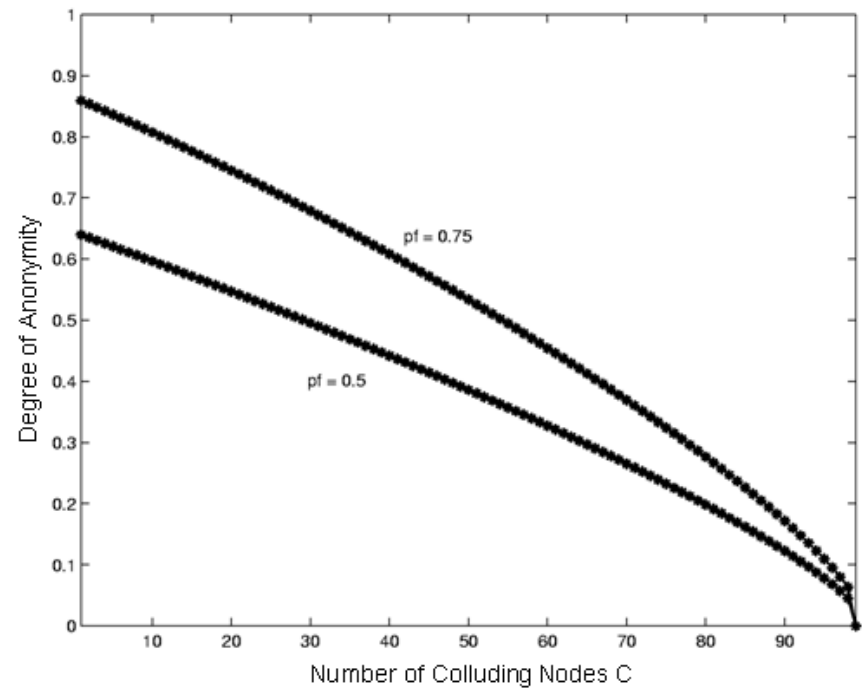
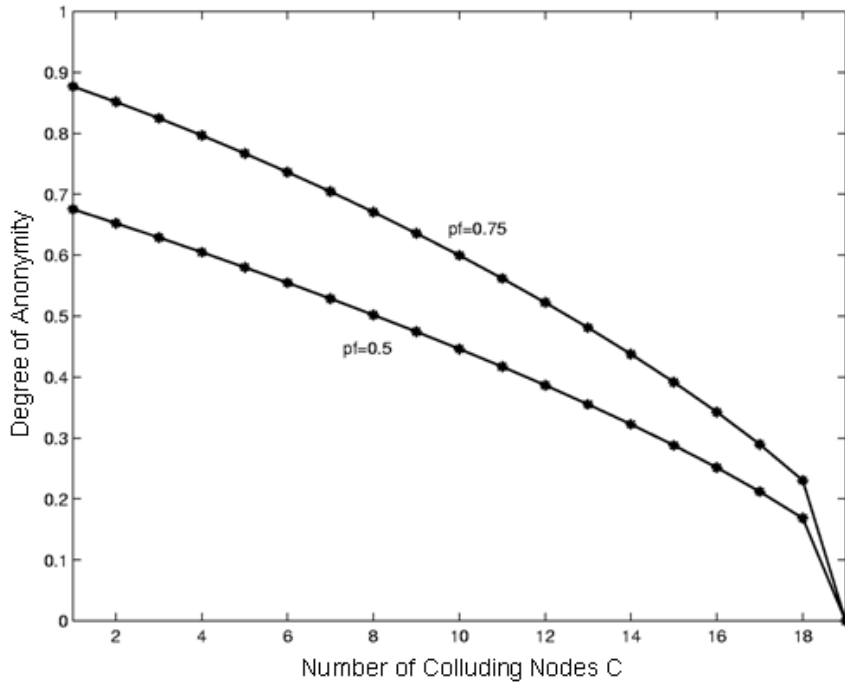
$N=7$

$C=2$

$$p_{C+1} = 1 - p_f \frac{N - C - 1}{N}$$

$$p_i = \frac{p_f}{N}, i = C + 2K \ N$$

Degree of Anonymity for Crowds



$$H = \frac{N - p_f(N - C - 1)}{N} \log_2 \left(\frac{N}{N - p_f(N - C - 1)} \right) + p_f \frac{N - C - 1}{N} \log_2 \left(\frac{N}{p_f} \right)$$

$$H_M = \log_2(N - C)$$

Information Theoretic Anonymity Metrics

- Degree of Anonymity
 - Systems with closed set of users (P2P networks)
 - Applied to Crowds
 - Tradeoff Anonymity/Scalability
- Effective Anonymity Set Size [SD02]
 - Entropy
 - Undetermined number of users (open systems)
 - Applied to Mixes

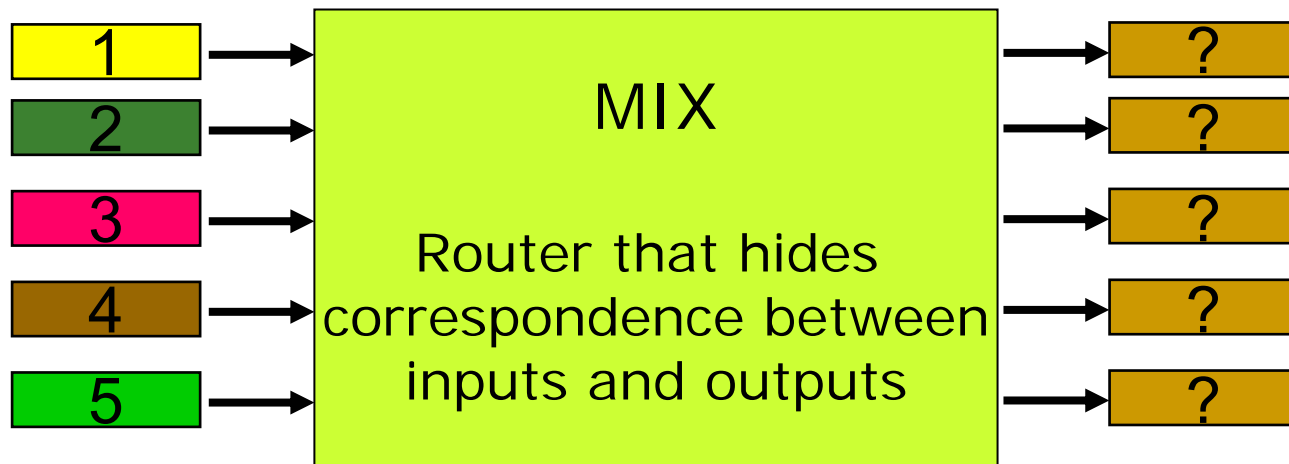
Contributions

- Quantification of anonymity
- Compare different anonymity systems
- Evaluate effectiveness of attacks
- Take into account statistical information obtained by the adversary

Outline

- Introduction to Anonymity
- Anonymity Metrics
- **Mixes**
- Passive Attacks - Outline
- Active Attacks
- Practical Evaluations - Outline
- Contributions and Open Questions

Concept of Mix



Functionality of Mixes

- Mixes modify
 - The appearance of messages
 - Encryption / Decryption
 - Padding / Compression
 - Substitution of information (e.g., IP)
 - The flow of messages
 - Reordering
 - Delaying - **Real-time requirements!**
 - Dummy traffic - **Cost of traffic!**

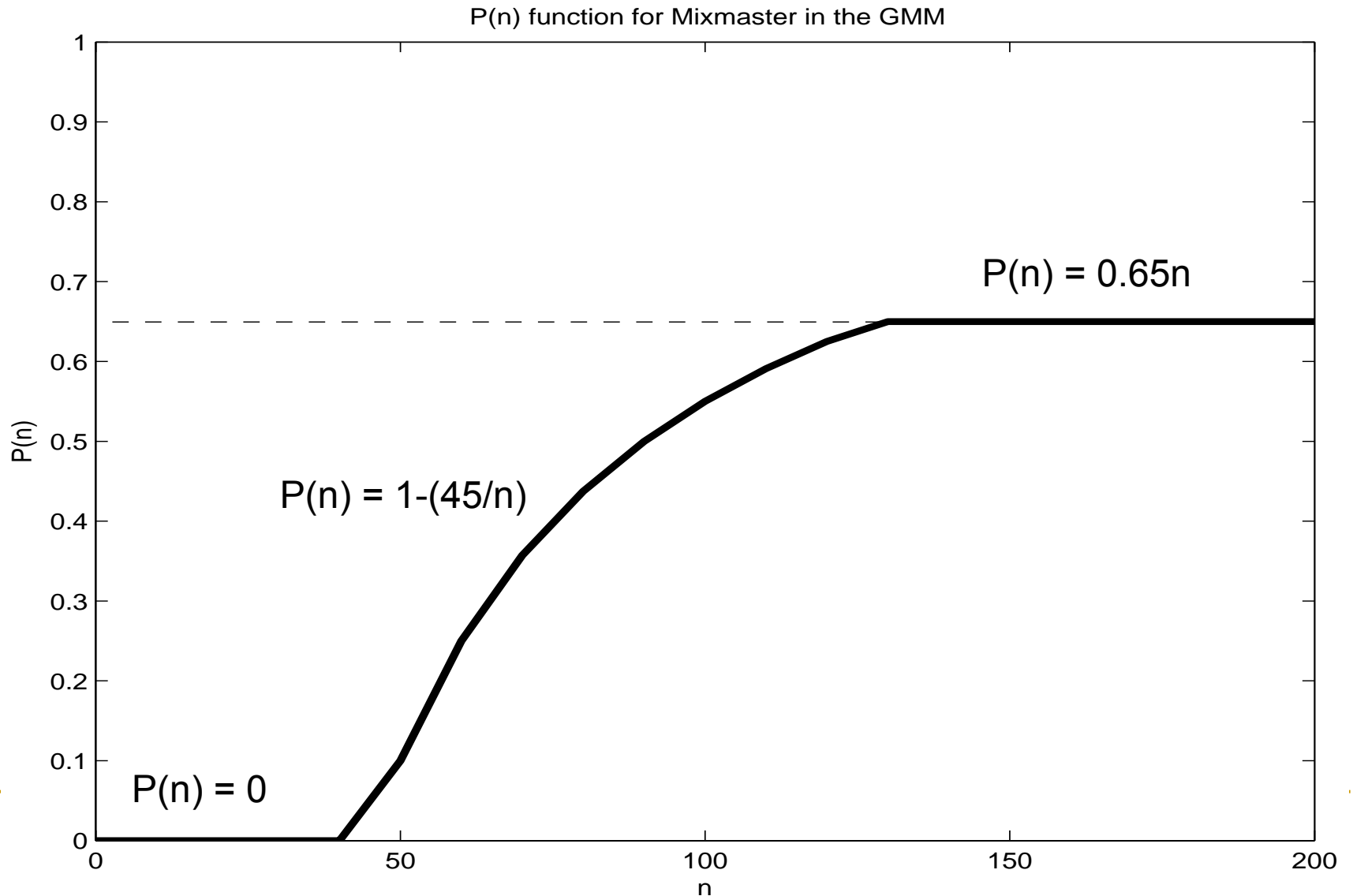
Pool Mixes

- Proposed by Chaum in 1981:
 1. Collect N inputs
 2. Shuffle
 3. Flush (Forward) } Round
- Pool selection algorithm
 - No pool / Static pool / Dynamic pool
 - Influences the performance and anonymity provided by the Mix
- Flushing condition
 - Time / Threshold
 - Deterministic / Random

Generalized Mix Model

- New way of thinking of mixes
- Mix represented when flushing condition is met
- $P(n)$ function: probability of a message from the pool of being forwarded as a function of the number n of messages in the mix
- All relevant data for anonymity is in the function
- Possibility of new complex pool selection algorithms (easy to design and implement)
- Possibility of randomization in selection (binomial mix)

Example: Mixmaster



Deterministic or Binomial?

■ Two ways of selecting messages

1. Send $n * P(n)$ messages, chosen uniformly at random
 - **Deterministic** number of messages
 - Randomness: **which** messages are selected?
2. Send every message with probability $P(n)$
 - **Binomial** number of messages, average $n * P(n)$
 - Randomness: **which** and **how many** messages are selected?
 - Influences the anonymity under certain circumstances (e.g., use of dummy traffic, N-1 attacks, ...)

Stop-and-Go Mix

- Proposed by Kesdogan in 1998
- Reordering strategy based on delaying
- $M/M/\infty$
- Delays generated by the user from an Exponential distribution
- Timestamping to prevent active attacks
- Based on the assumption of Poisson incoming traffic, it provides anonymity estimates

Contributions

- Taxonomy mixes
- Generalized Mix Model (GMM)
 - Simple model that captures all the relevant information for anonymity
 - Possibility of new functions
- Binomial mixes
 - Added randomness
 - Additional cost: just a few extra calls to Random Number Generator
 - Hides information from attacker
 - More robust towards certain attacks

Outline

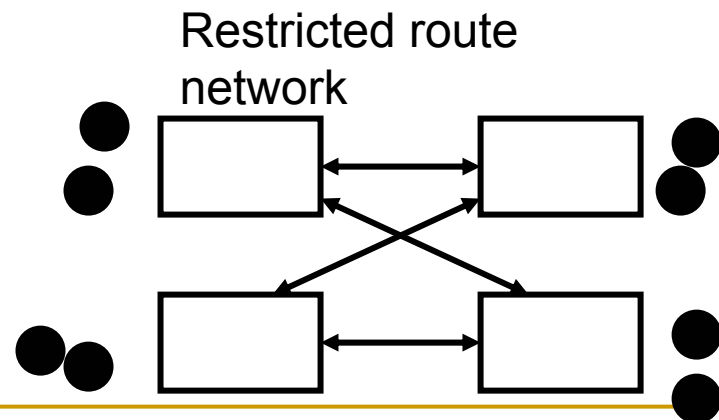
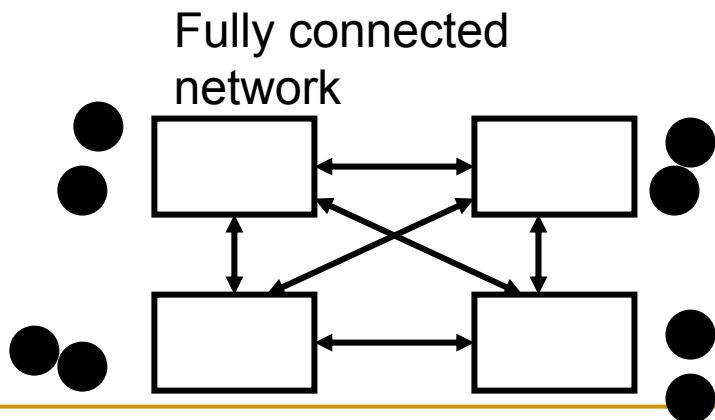
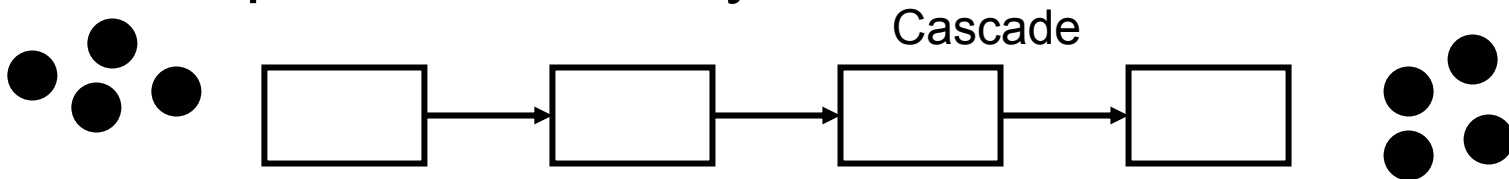
- Introduction to Anonymity
- Anonymity Metrics
- Mixes
- **Passive Attacks - Outline**
- Active Attacks
- Practical Evaluations - Outline
- Contributions and Open Questions

Passive Attack on Mixes

- We have computed the probability distributions and applied information theoretic anonymity metrics to:
 - Pool Mixes
 - Various $P(n)$ functions (GMM)
 - Timed / Threshold
 - Deterministic / Binomial
 - Continuous Mixes (no assumptions on traffic pattern)
 - Exponential Delays
 - Uniform Delays

Mix Networks

- Mixes are combined in networks in order to
 - Distribute trust
 - Improve availability



Dummy Traffic

- Fake messages introduced to confuse the attacker
- Created and discarded by mixes
- Very useful in low traffic conditions
- Inserted at output or in pool (pool mixes)
- Dummies improve the anonymity by making more difficult the traffic analysis
 - Increase of traffic
 - Increase of possible paths in network
 - **Impact in nodes that generate/discard the dummies**

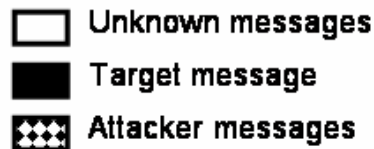
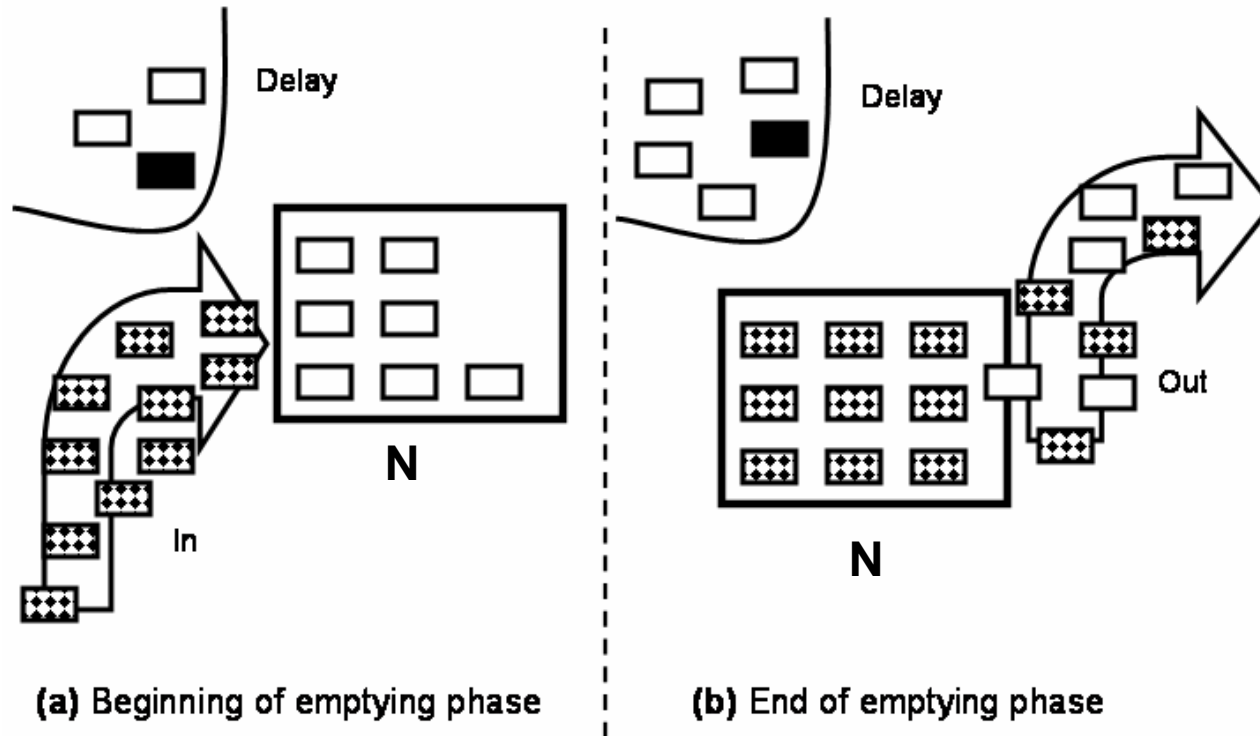
Passive Attacks: Pool Mixes that Generate Dummy Traffic

- First study on the impact of dummy traffic on the anonymity of mixes that generate / discard dummies
- Intuitive extension of the metric leads to meaningless results
- Explain how to correctly compute anonymity
 - Dummies inserted at the Output
 - Dummies inserted in the Pool
- Dummies inserted (deleted) by mix contribute only to recipient (sender) anonymity

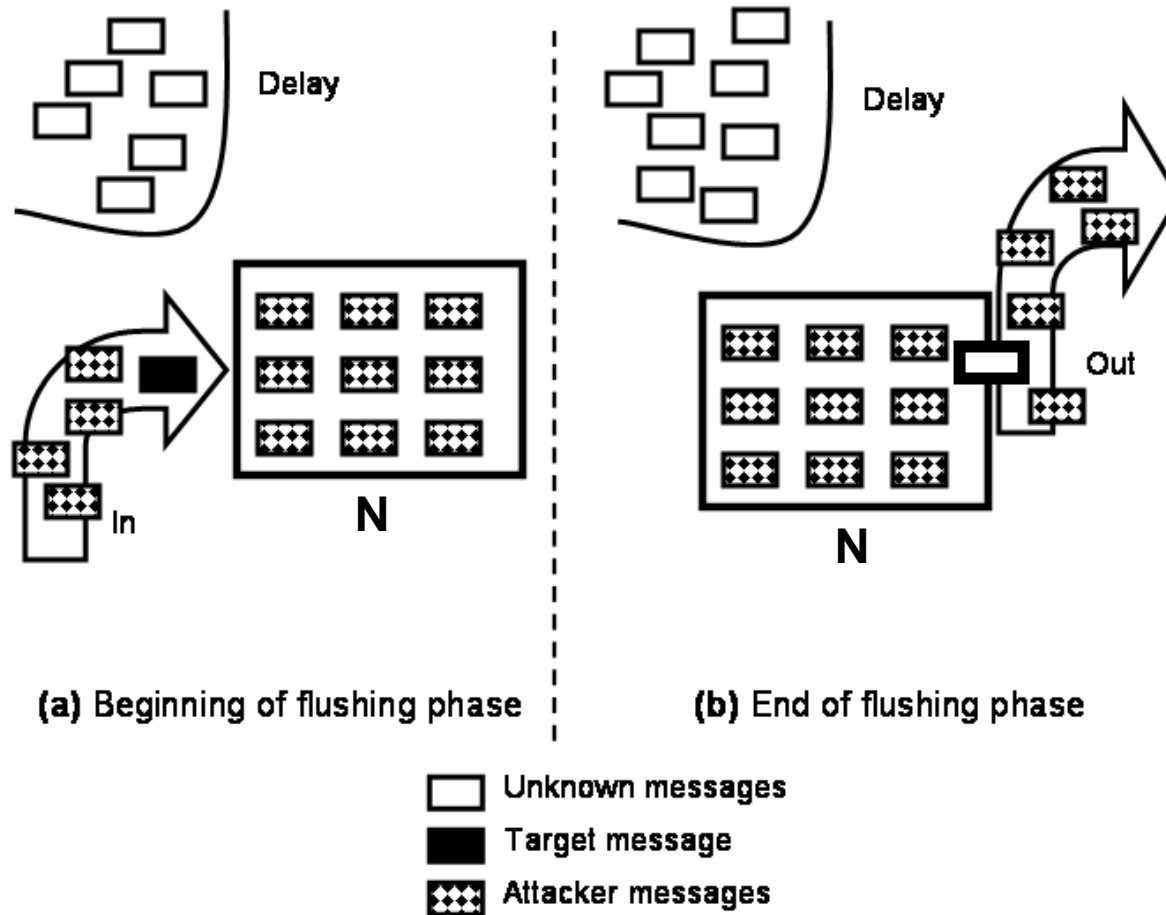
Outline

- Introduction to Anonymity
- Anonymity Metrics
- Mixes
- Passive Attacks - Outline
- **Active Attacks**
- Practical Evaluations - Outline
- Contributions and Open Questions

N-1 attack: Emptying Phase



N-1 Attack: Flushing Phase



Effort of the Attacker

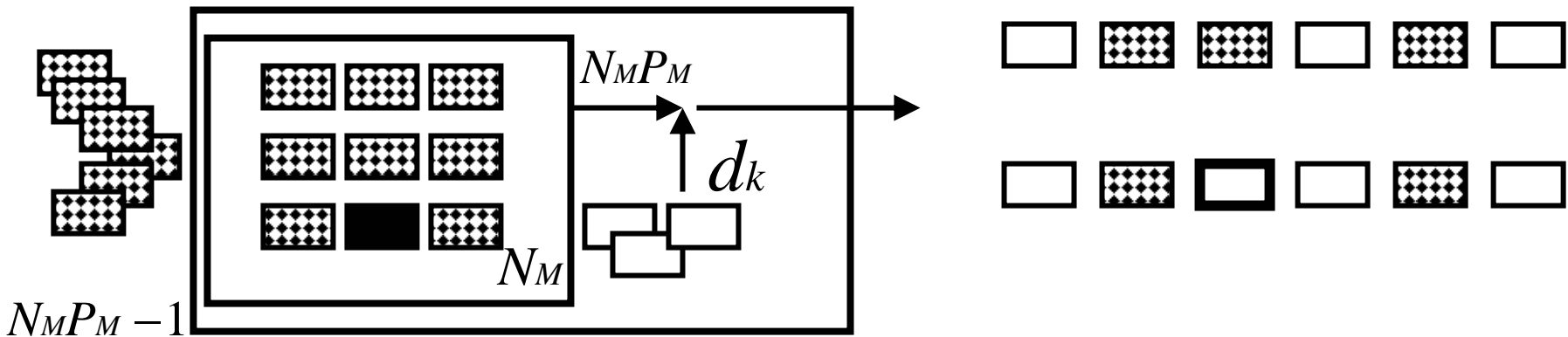
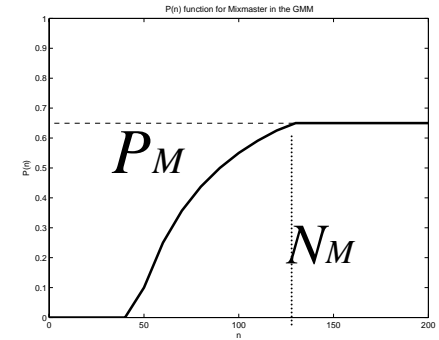
- T_a : Time required to deploy the attack
- R_a : Number of rounds required for the attack
- N_u : Number of unknown messages delayed by the attacker
- N_a : Number of messages generated by the attacker
- T_p : Observation time needed to establish the state of the mix
- R_p : Observation number of rounds
- H : Remaining anonymity

Mixes Evaluated

- Timed Pool Mixes
- Threshold Pool Mixes - fastest
- Binomial Pool Mixes – two flavors
- Continuous Mixes - delaying
- **Deterministic and binomial** pool mixes with **dummy traffic (only flushing phase)**
 - **Inserted at output**
 - **Inserted in pool**

Flushing Phase on Pool Mix with Dummies at output

- Function $P(n)$, max P_M at N_M
- Mix inserts at output of round k d_k dummies



Flushing Phase with Dummies Inserted at the Output: Deterministic Pool Mix

$$R_a(\text{flush}) = \frac{1}{P_M}$$

$$T_a(\text{flush}) = T \cdot R_a(\text{flush}) = \frac{T}{P_M}$$

$$N_a(\text{flush}) = P_M \cdot N_M \cdot R_a(\text{flush}) = N_M$$

$$N_u(\text{flush}) = m_u \cdot T_a(\text{flush}) = m_u \cdot T \cdot \frac{1}{P_M}$$

$$H = -\sum_{j=1}^{d_k+1} \frac{1}{d_k+1} \log_2\left(\frac{1}{d_k+1}\right) = \log_2(d_k+1)$$

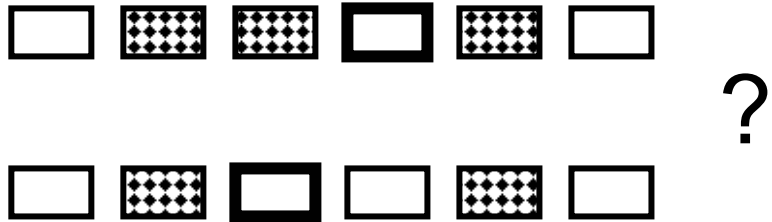
$$\Pr(R_a(\text{flush}) = 1) = P_M$$

$$\Pr(R_a(\text{flush}) = i) = P_M(1 - P_M)^{i-1}$$

$$E(R_a(\text{flush})) = \frac{1}{P_M}$$

$$P_M = 0.6 \rightarrow R_a(\text{flush}) = 1.6$$

Binomial Mix



$$(1 - P_M)^{Ra(\text{flush})} \leq \varepsilon$$

$$\Pr(O_{i,j}) = \frac{P_M (1 - P_M)^{i-1}}{u_i}, \quad j = 1..u_i$$

$$H = - \sum_{i=1}^{Ra(\text{flush})} \sum_{j=1}^{u_i} \Pr(O_{i,j}) \log_2(\Pr(O_{i,j}))$$

$$H = - \sum_{i=1}^{Ra(\text{flush})} P_M (1 - P_M)^{i-1} \log_2\left(\frac{P_M (1 - P_M)^{i-1}}{u_i}\right)$$

$$P_M = 0.6$$

$$\varepsilon = 0.01$$

$$Ra(\text{flush}) = 6$$

Contributions

- Define set of parameters that characterize the effort of deploying an N-1 attack
- Compute effort for several pool mixes and continuous mix
- Method to find the number of messages contained in the pool of a binomial mix
- Compute remaining anonymity
- Dummy traffic:
 - Impact on effort
 - Impact on remaining anonymity

Outline

- Introduction to Anonymity
- Anonymity Metrics
- Mixes
- Passive Attacks - Outline
- Active Attacks
- **Practical Evaluations - Outline**
- Contributions and Open Questions

Mixmaster Remailer Network

- Operative since 1995
- Anonymous email service
- Mixmaster Nodes
- Reliable Nodes
- Collection of real inputs (node since 2000)
- Simulation of Mixmaster and Reliable
- Analysis of anonymity
 - Sender/Recipient Anonymity
 - Correlation with Delay and Traffic Load

Contributions

- Evaluation of working implementations
- Analysis of input traffic
 - Not necessarily Poisson!
- Analysis of Mixmaster
 - Minimum anonymity provided, independently of traffic
- Analysis of Reliable
 - No anonymity if traffic too low
 - Delays not too large
- Evaluation of implementation: sources of randomness, crypto libraries, usability, host server integrity, UI, Documentation

Conclusions

- Degree of Anonymity (Information Theoretic Anonymity Metric)
- Generalized Mix Model
- Binomial Mixes
- Analysis of Theoretic Mixes
 - Anonymity Towards Passive Adversary
 - N-1 attack: Effort of Active Adversary and Remaining Anonymity
 - Analysis of the impact of Dummy Traffic (Passive and Active)
- Analysis of Implemented Anonymous Email Service

Open Questions

- Other Information Theoretic Metrics
 - Mutual Information?
 - Min-Entropy?
- Apply Anonymity Metrics to Mix Networks
- Find Optimal Mixing Strategies
- Find Optimal Dummy Strategies
- Solutions for Low-Latency Anonymous Communication?

Publication List

- 1 International Journal
- 5 Lecture Notes in Computer Science (LNCS)
- 2 International Conferences / Workshops
- 1 National Journal
- 1 National Conference