The Wisdom of Crowds: attacks and optimal constructions

Claudia Diaz
K.U.Leuven, ESAT-COSIC

ZISC Information Security Colloquium
December 1st, 2009

Acks for the slides

George Danezis[1], Claudia Diaz[2], Emilia Käsper[2], and Carmela Troncoso[2]
[1]Microsoft Research Cambridge [2]K.U.Leuven, ESAT-COSIC

(paper published in ESORICS 2009)

Outline

- Introduction: traffic analysis and anonymous communications
- Anonymous Peer-to-Peer Routing via Crowds
  - The Crowds scheme (1998)
  - Security of Crowds
- The Always Down-or-Up Scheme (ESORICS '08)
  - The ADU routing mechanism
  - Traffic analysis of ADU
- Optimality of Crowds
  - A general model for message-passing
  - Optimality of Crowds in the model
  - Performance trade-offs

Traffic analysis

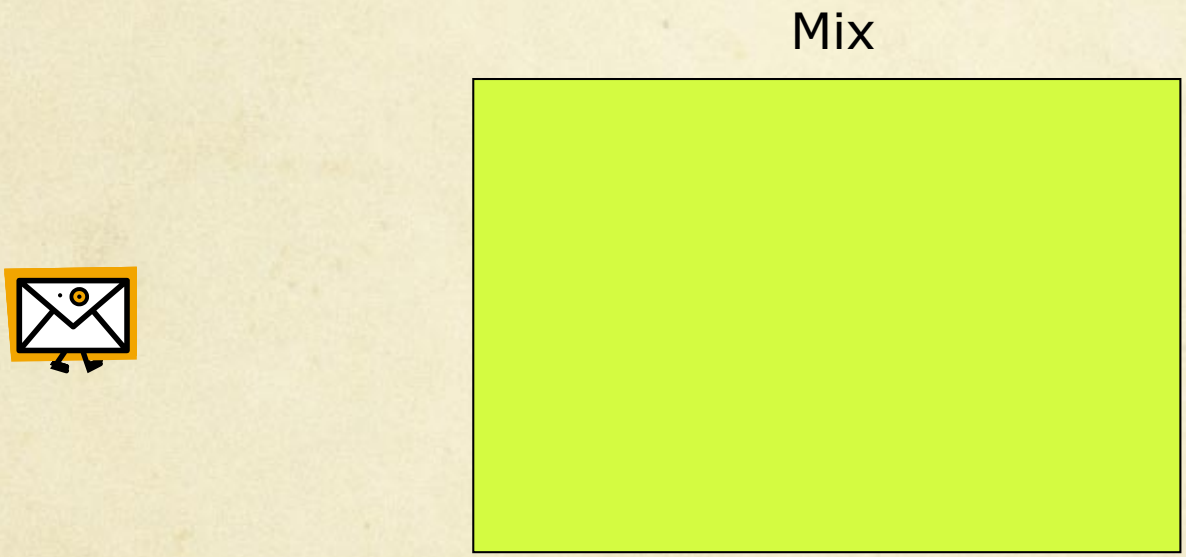
- Traffic data: source, destination, timing, volume, etc
- Even if communication is encrypted, traffic data can reveal a lot of information
- Examples from WW II (signals intelligence):
  - Traffic analysis was used by the British at Bletchley Park to assess the size of Germany's air-force
  - Dummy traffic to conceal D-Day action
  - Identifying people by their typing
- Amateur plane-spotters revealed the CIA's `extraordinary rendition' programme

Traffic analysis

- Traffic data has less volume than content:
  - Coarser, but highly valuable information
  - Can be used to select targets for more intensive surveillance
  - Formats that are easy to process for machines
  - Harder to conceal
- Diffie and Landau, in their book on wiretapping: “traffic analysis, not cryptanalysis, is the backbone of communications intelligence“
- Anonymity at the communication layer is assumed as building block for other privacy protection technologies

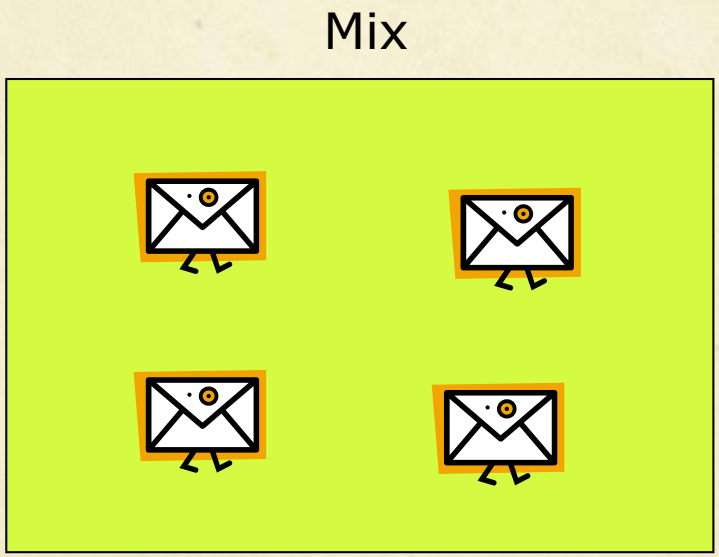
Chaumian Mix (Chaum 1982)

- Message-based, high latency
- Phase 1: collect inputs
- Parameter T (threshold): T=4 in example



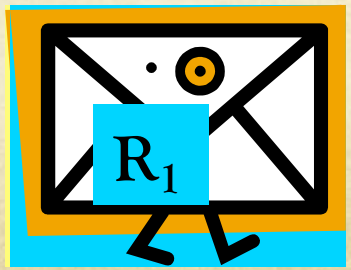
Chaumian Mix (Chaum 1982)

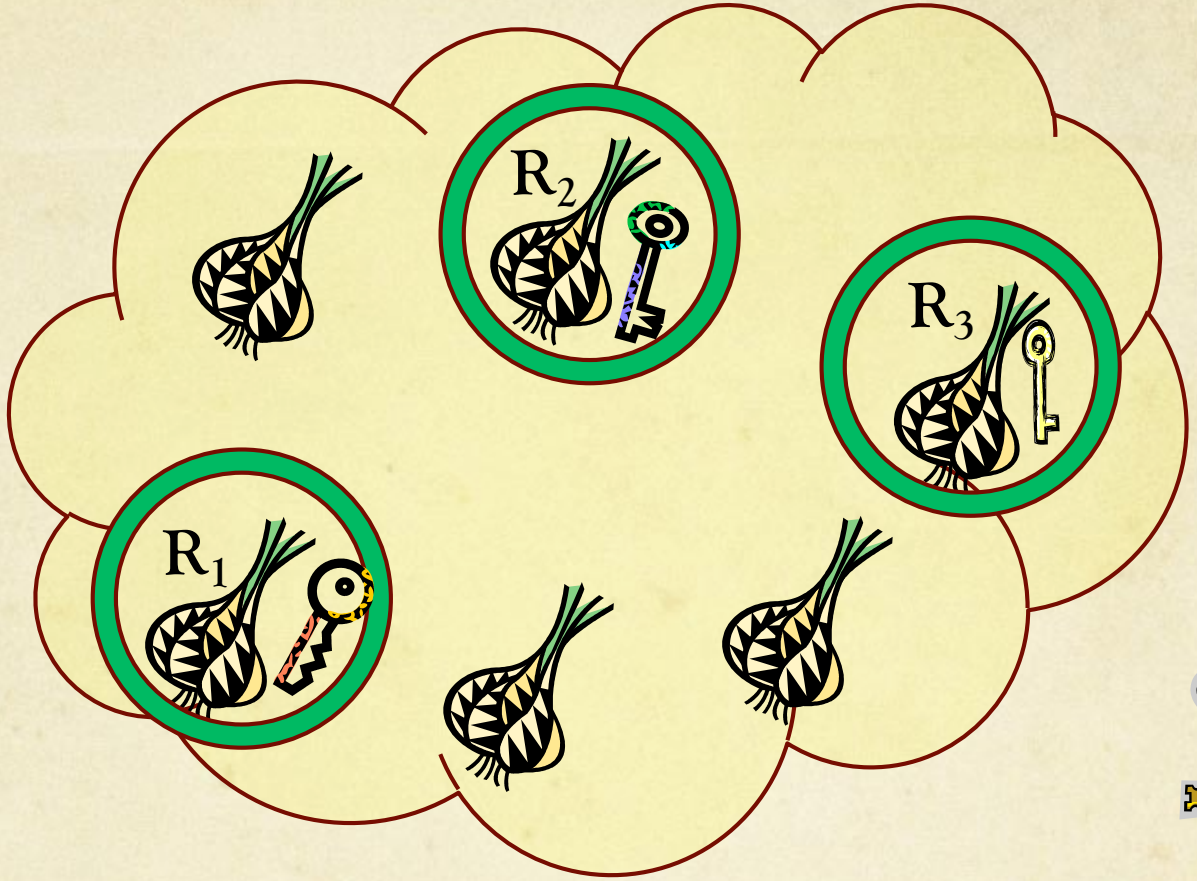
- Phase 2: mix and flush



Onion Routing (1998): onion encryption

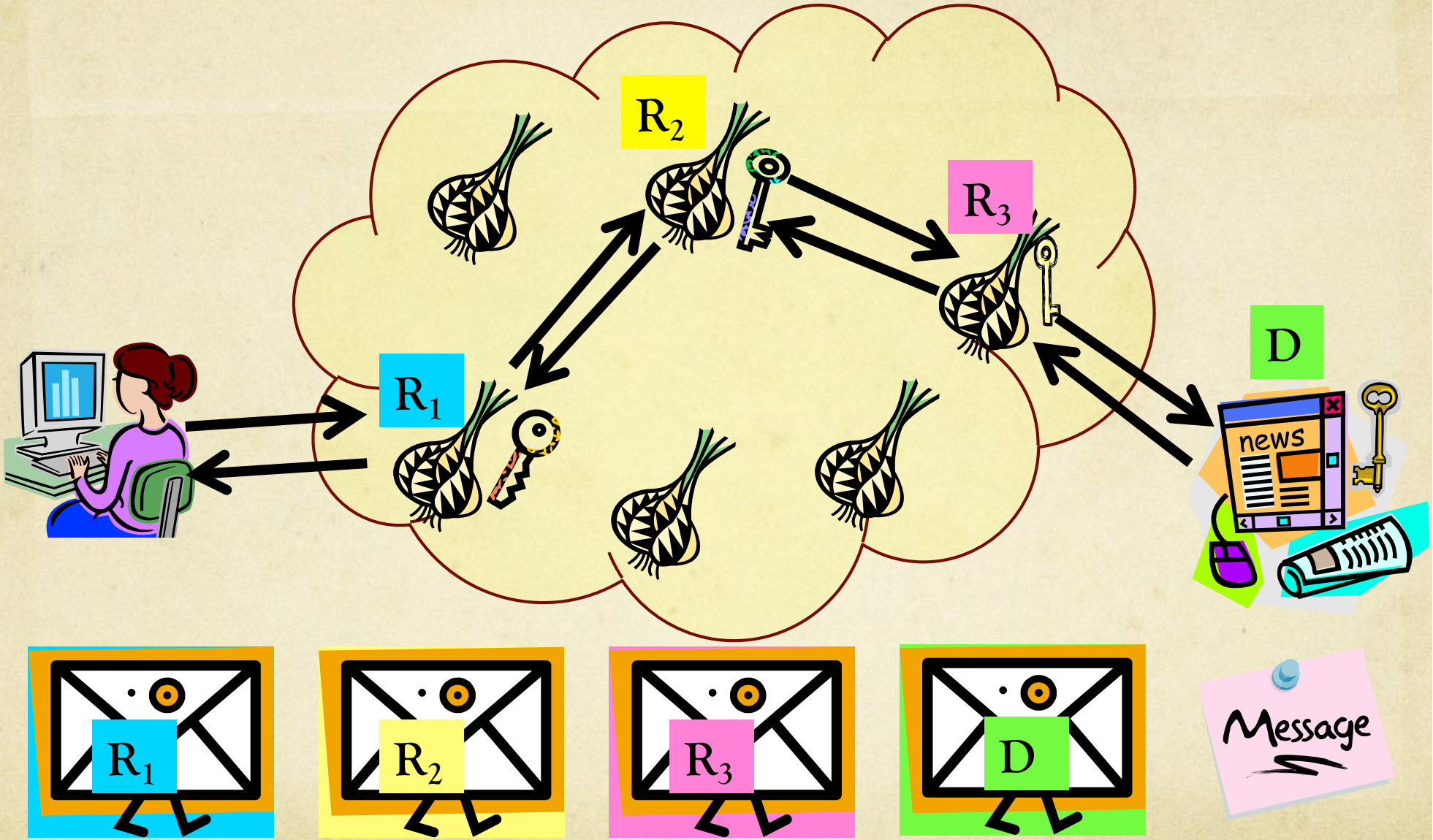
Source routed



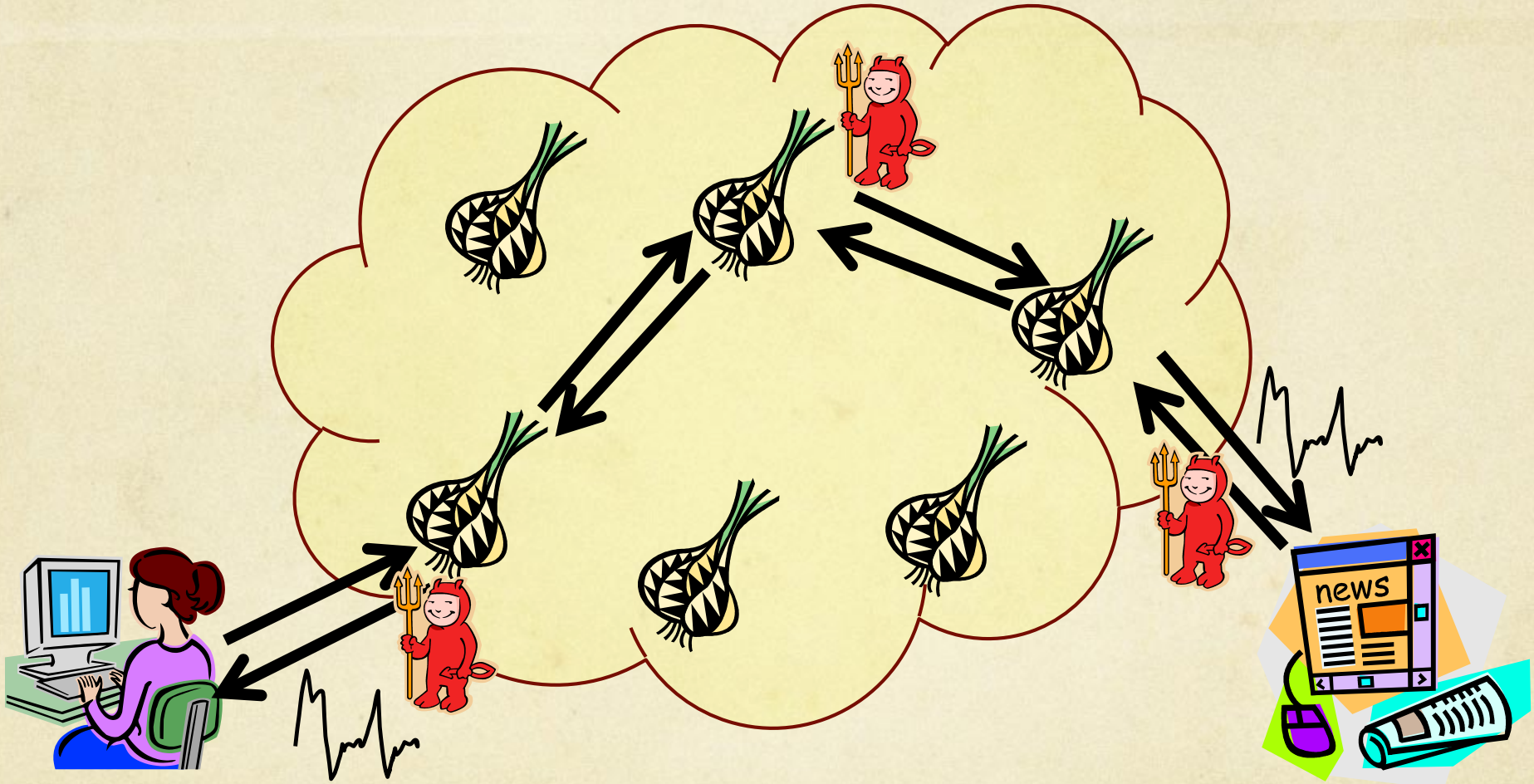


D

Onion routing

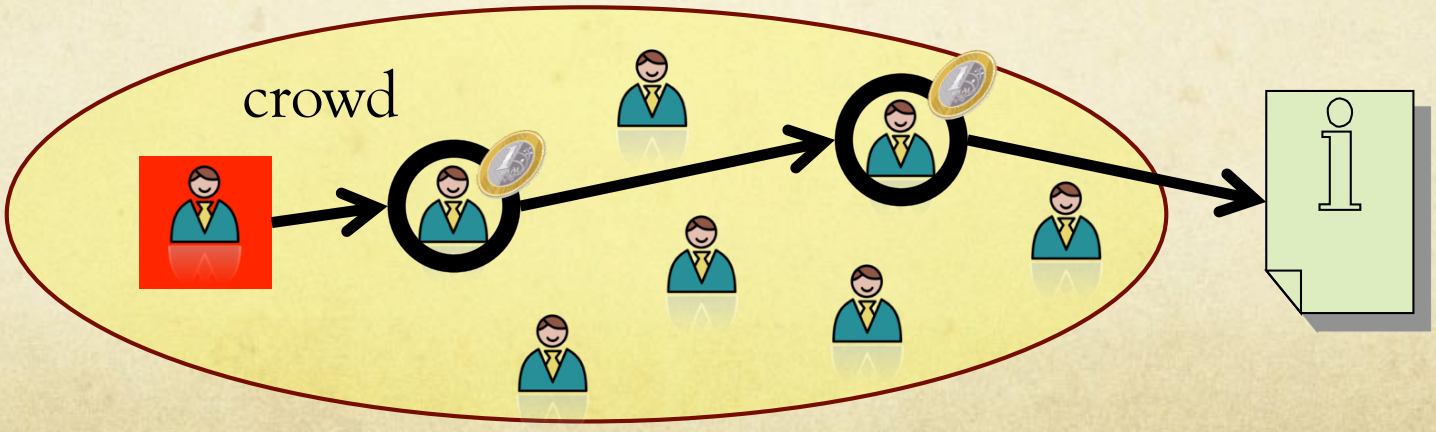


TOR - adversary model



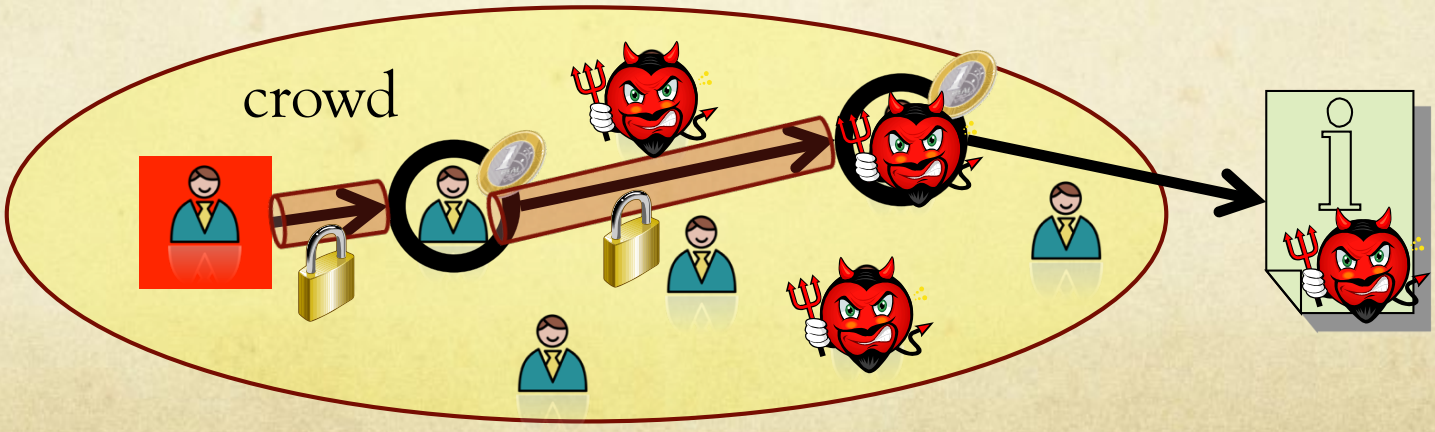
Crowds (Reiter, Rubin 1998)

- Anonymity for web browsing
- Group of users form a “crowd”
- Initiator chooses a random member of the crowd and forwards the web request to her
- The recipient of the request flips a biased coin and forwards the request to the end server with probability p and to another member with probability 1-p
- A tunnel is established between the initiator and the exit crowd member (static paths)



Crowds (Reiter, Rubin 1998)

- Communication between members is encrypted with symmetric keys
  - BUT: all members can see the request in clear
- Adversary model:
  - Assumed adversary cannot control all links
  - Instead, the adversary controls a subset of the crowd and/or the end server
- Probability that predecessor/exit is the initiator or just a forwarder



Anonymity of Crowds wrt the destination

- The message always travels at least one hop in the crowd
- The end server receives the message from a random crowd member
- The probability that the last node before the destination is the sender of the message is 1/N in a crowd of size N
- The a priori probability is also 1/N
  - The end server gains no additional information by observing the message
- Thus, Crowds provides optimal anonymity against the destination

Anonymity of Crowds wrt corrupt nodes

- Assume an adversarial node receives a message
- The adversary has to decide whether the previous node is the sender of the message
- In other words, he has to decide whether he is the first node on the path
- In a crowd with parameter p and fraction of corrupt nodes f , this probability is:

$$\Pr[\text{previous} = \text{sender} | \text{message}] = 1 - (1 - p)(1 - f)$$

- E.g. p = 0.33, f = 0.1: 40% certainty that the previous node is the sender.

Improving upon Crowds

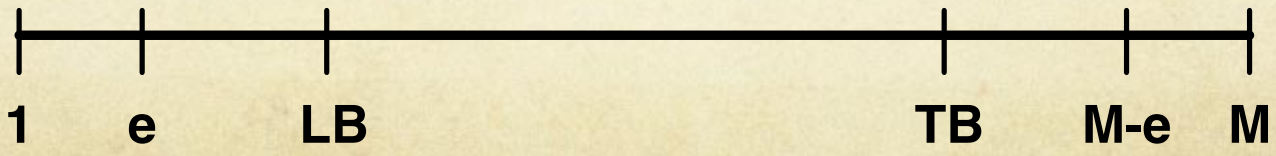
- The sender can be determined with certainty

$$1 - (1 - p)(1 - f)$$

- We cannot control the number of corrupt nodes f
- In order to increase anonymity, we must choose a smaller parameter p
- Decreasing p increases the mean path length
- Question: Are there alternative message-passing algorithms that provide better latency without a compromise in anonymity?

The Always Down-or-Up (ADU) scheme [ESORICS '08]

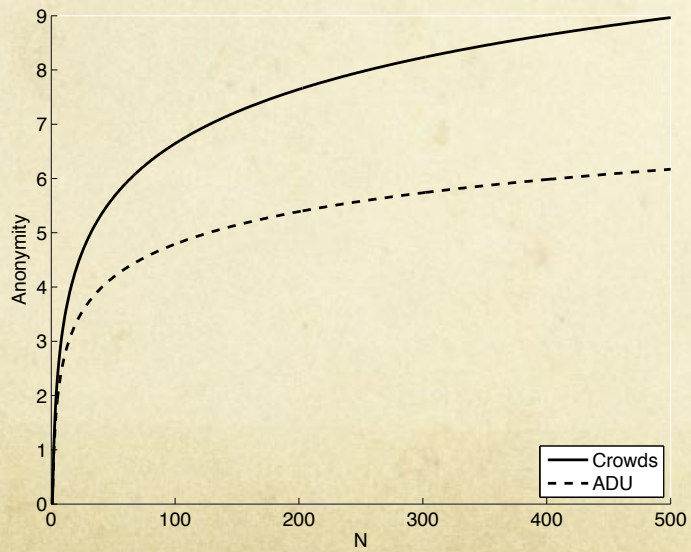
- The sender chooses an integer u_0 in the interval [1,M]
- If $u_0 \leq e$ or $u_0 \geq M-e$ send message to end destination
- If $u_0 \leq LB$ ($u_0 \geq TB$) choose mode AD (AU)
- Else choose mode randomly
- Forward u_0 and mode AD/AU to a random node
- In AD mode: each subsequent node moves down in the interval by choosing $u_{i+1} \in [1,u_i)$. The message is sent to destination when $u_i \leq e$.
- In AU mode: move up analogously

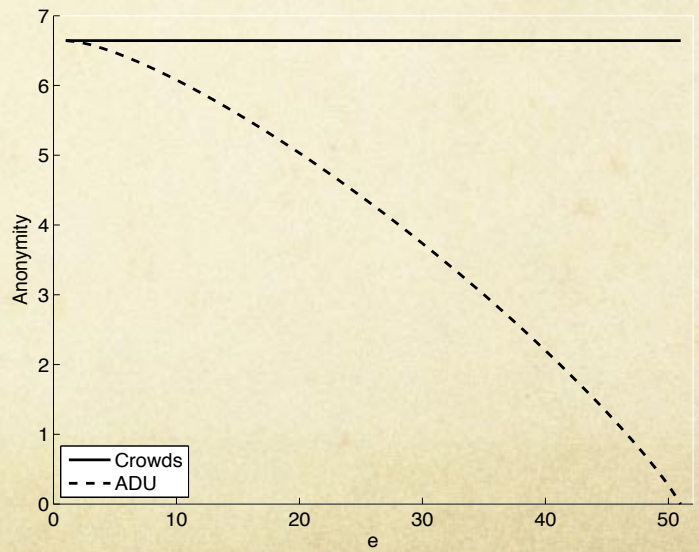


Traffic analysis of ADU at the destination

- A fraction of messages are sent directly to the destination
- A message received at the destination is more likely to come from the true sender than any other member of the crowd

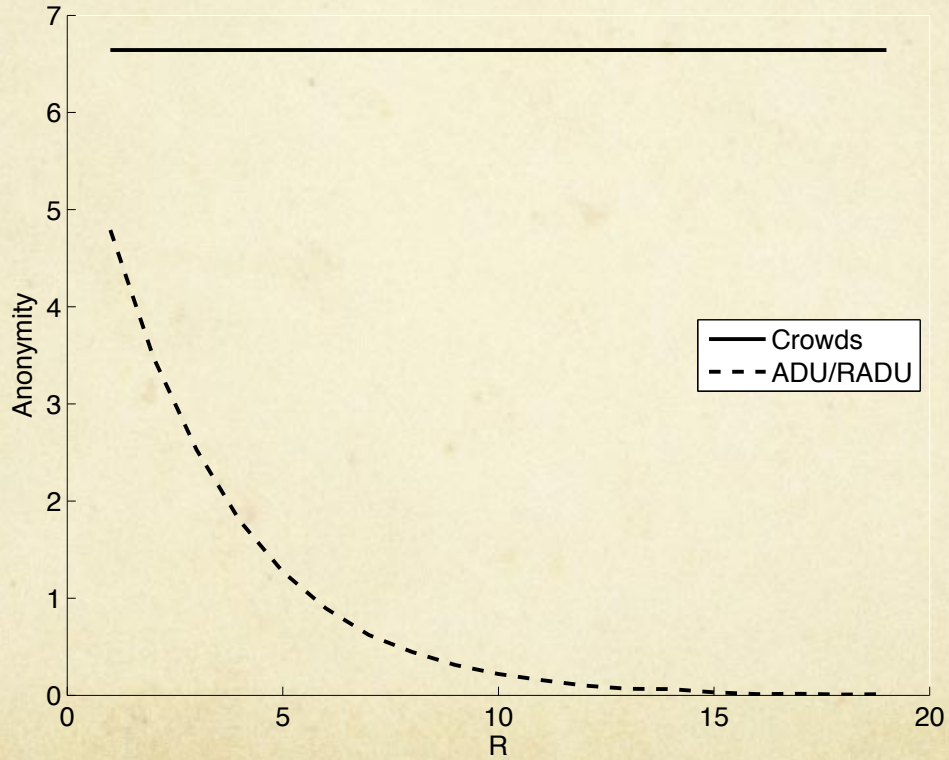
$$\Pr[n_i \mid n_i] = \frac{2e}{M} + (1 - \frac{2e}{M})\frac{1}{N} \qquad \Pr[n_x \mid n_i, x \neq i] = (1 - \frac{2e}{M})\frac{1}{N}$$





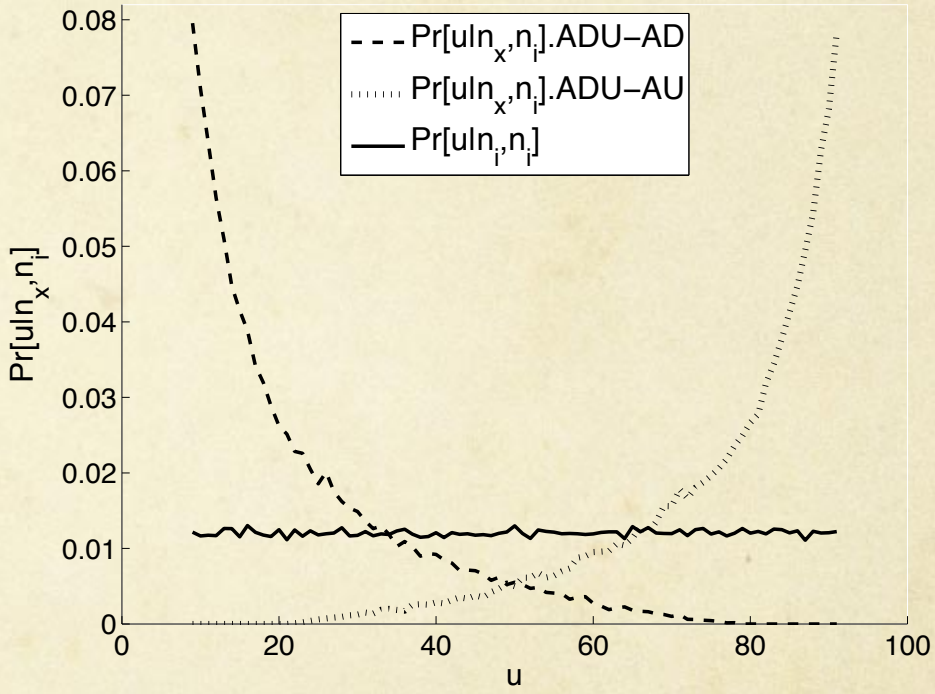
Multiple requests

- Anonymity decreases further as multiple requests are made



Traffic analysis of ADU in the crowd

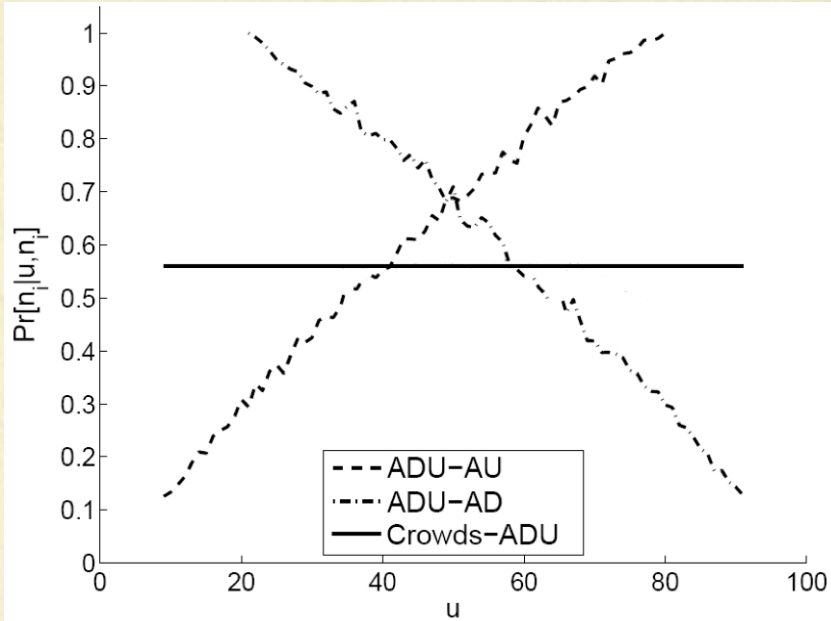
- Varying the mode Always-Down vs Always-Up has no security merit: the mode is fixed and the adversary knows it
  - We analyzed a randomized variant: its security is in between ADU and Crowds
- The value u_i follows different distributions when
  - predecessor = initiator
  - predecessor ≠ initiator



Traffic analysis of ADU in the crowd

- The adversary uses the parameter u to better decide if the predecessor is the initiator or not

$$\Pr[n_i \mid u, n_x] = \frac{\Pr[u \mid n_x, n_i]\Pr[n_x \mid n_i]\Pr[n_i]}{\sum_j \Pr[u \mid n_x, n_j]\Pr[n_x \mid n_j]\Pr[n_j]}$$



A general model for message-passing in a crowd

- Each node sees the message body, the destination, and some arbitrary routing information
- Each node must have sufficient routing information to decide whether to pass the message on or send it to the destination
- A corrupt node can simulate routing by forwarding the message to itself and thus necessarily learns the number of remaining hops—the time-to-live (TTL) of the message
- On the other hand, the TTL is sufficient to route correctly
- All additional information is redundant and can only harm the security of the system

D-Crowds for arbitrary distributions

- The sender draws a time-to-live TTL from some distribution D
- She then forwards the message along with the TTL to a randomly chosen crowd member
- Each subsequent node
  - Forwards the message to the destination if TTL=0;
  - Forwards the message and the new time-to-live TTL=TTL-1 to a random node otherwise.
- The D-Crowds model captures all message-passing algorithms that leak minimal information
- Crowds is equivalent to D-Crowds with a geometric distribution $D \approx Geom_p$.

Measuring anonymity in the crowd

- Worst-case security: We measure the maximum probability of determining the sender over all messages
- Average-case security guarantee is not enough
  - We do not know the cost of a single compromise
  - Each user cares about her own message
- For meaningful comparison, we always require perfect security against the end server
  - In a trivial system where all messages are sent directly to the server, the user has perfect anonymity in the Crowd.

The optimality of Crowds

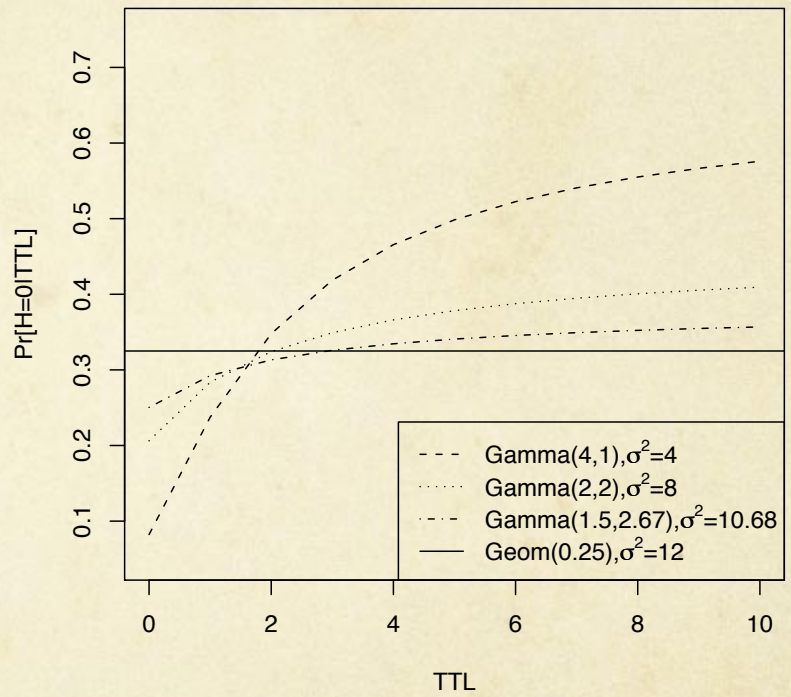
- Let Adv^f (D) be the maximum probability with which the sender can be determined, for distribution D

$$Adv^f(D) = \max_{ttl} \Pr[\text{predecessor} = \text{initiator} \mid TTL = ttl]$$

- Theorem
  - For an arbitrary distribution D(l) over path lengths, if for all f , 0 < f < 1,
  - Adv^f (D) ≤ Adv^f ($Geom_p$),
  - Then E(D) ≥ E($Geom_p$).
- Thus, Crowds provides optimal anonymity for any given mean message path length.

Trade-Off: path length variance vs anonymity

- Non-geometric distributions provide suboptimal anonymity
- Performance trade-off: distributions with weaker anonymity may offer lower variance in path length



Conclusions

- The TTL-based D-Crowds model captures all “sensible” message-passing algorithms
- The original Crowds provides optimal anonymity under this model
- Our main result: if two schemes have equal mean path length, then the anonymity guarantees provided by Crowds are stronger
- The lesson: When designing a scheme, be suspicious of free lunches. The less latency and variance in latency, the less anonymity a system is likely to provide.

Thank you!

http://homes.esat.kuleuven.be/~cdiaz/
claudia.diaz@esat.kuleuven.be