The Wisdom of Crowds: attacks and optimal constructions

George Danezis[1] Claudia Diaz[2] Emilia Käsper[2]
Carmela Troncoso[2]

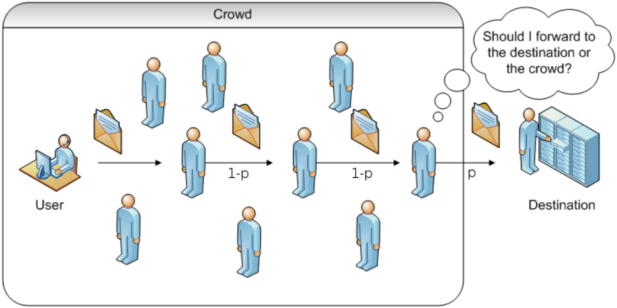
[1]Microsoft Research Cambridge

[2]Katholieke Universiteit Leuven, ESAT-COSIC

ESORICS 2009
Saint-Malo, September 2009

The Crowds scheme (1998)

- The sender uses a P2P network to communicate anonymously with a destination
- Each intermediate node flips a biased coin to decide whether to forward the message in the crowd or to the destination



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 $\frac{1}{N}$ in a crowd of size N.
- The *a priori* probability is also $\frac{1}{N}$— 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

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

- 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

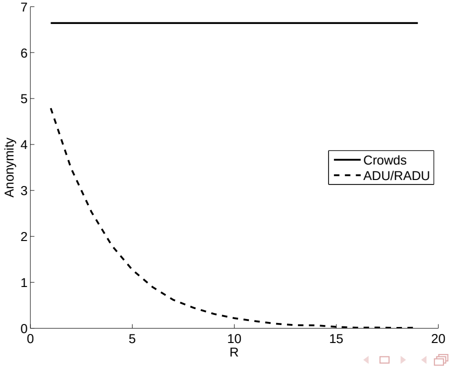
Are there alternative message-passing algorithms that provide better latency without a compromise in anonymity?

ADU: the Always Down-or-Up 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

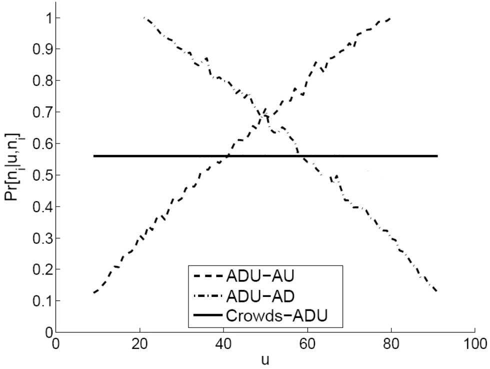
Traffic analysis of ADU at the destination

- 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
- 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
- The value u_i leaks information on how long the message has travelled in the crowd



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: I will not send out a vulnerable message!
  - Compare with cryptography: I want ***my*** RSA key to be strong.
- 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 $\mathrm{Adv}^f(D)$ be the maximum probability with which the sender can be determined, for distribution D.

Theorem

For an arbitrary distribution $D(l)$ over path lengths, if for all f, $0 < f < 1$,

$$\mathrm{Adv}^f(D) \leq \mathrm{Adv}^f(Geom_p),$$

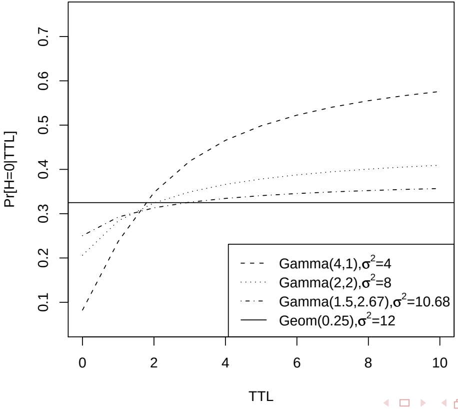
then

$$\mathbb{E}(D) \geq \mathbb{E}(Geom_p).$$

- 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.