Universität zu Köln

Mathematisches Institut

Discourse

Anonymity Networks

Laslo Hunhold

In the lecture

'Information Theory and Statistical Physics'

Prof. Dr. Johannes Berg

4th July 2017

Contents

1 Introduction

In times of increasing corporate and state surveillance and exchange of critical information by digital means, the need for secure information storage and transmission has risen dramatically. One important concept is *deniability*, which plays a major role in data encryption and describes the indistinguishability of encrypted data from random noise. This discourse takes a look at the less famous but similar concept of *deniable communication*, which seeks to hide the sender of a message using a statistical process for message forwarding within a so-called *anonymity network*.

We use statistical and information theoretical means to formulate the general theory and derive properties of the anonymity network 'Crowds' (see [RR98]). Our goal is to prove that within certain bounds, a given anonymity network allows deniable communication even when a certain number of peers is compromised. Known attacks like the 'predecessor attack' (see [WALS04]) and 'intersection attack' (see [DS05]) will not be discussed given the scope of this document.

2 Anonymity Networks

There are multiple concepts of attackers on an anonymity network and we make the following assumptions for a general network that provides the biggest possible liberty in a decentralized context: We are a '*passive attacker*', as messages can not be altered meaningfully given they are encrypted by design. Nodes can be compromised ('*internal attacker*'), as each node joining the network has the same importance regardless of it being good or evil, and we are a '*global attacker*', as the passing of messages happens across all nodes.

Assume that for a given network of $N \in \mathbb{N}$ nodes $n_1, \ldots, n_N$ we control $0 \leq C < N$ corrupt nodes (see figure 2.1). If a message is passed to one of our corrupt nodes, we are trying to find out who the most probable sender was. It has not been established

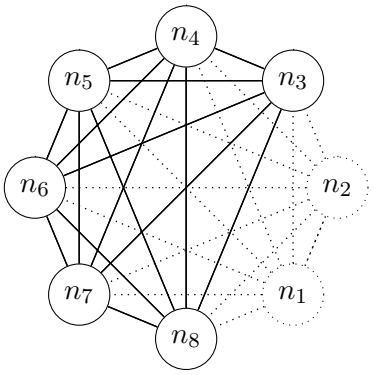

Figure 2.1: Anonymity network for $(N, C) = (8, 2)$.

here yet, but the passing of messages is based on a statistical process and the message itself gives no information on which node it originated from. We can however assume that each node has the same weight and that our corrupt nodes are indistinguishable from good ones by showing no suspicious behaviour.

We define a discrete random variable X that allows us to express the probability of a node being the sender as

$$\mathbb{P}(X = n_i) := p_i.$$

Given we don't initiate any messages from our collaborating nodes, we know of them to have zero probability.

2.1 Degree of Anonymity

The proposal to define degrees of anonymity can already be found in [RR98], though not formalized. The idea to use the SHANNON entropy for this purpose was first brought up in [DSCP03] and we will use the definition given in [SD03].

Regardless of how the messages are passed, we can already define a *degree of anonymity* using X. To do that, we take a look at the information entropy of the best and worst case scenario. In the ideal case with $X = \overline{X}$, all good nodes have the uniform distribution $p_i = \frac{1}{N-C}$, which yields

$$\overline{H} := H(\overline{X}) = -\sum_{i=1}^{N} p_i \cdot \ln(p_i) = -(N-C) \cdot \frac{1}{N-C} \cdot \ln\left(\frac{1}{N-C}\right) = \ln(N-C).$$

In the worst case with $X = \underline{X}$, we already know where the message originated from and we have $p_i = 0$ except for one node, the message initiator with probability 1, which yields

$$\underline{H} := H(\underline{X}) = -\sum_{i=1}^{N} p_i \cdot \ln(p_i) = 1 \cdot \ln(1) = 0.$$

Based on this observation, we define a quantity which is 1 for $\overline{X}$ ('100% anonymity') and 0 for $\underline{X}$ ('0% anonymity') as follows:

Definition 2.1 (degree of anonymity, [SD03])**.**

$$d(X) := 1 - \frac{\overline{H} - H(X)}{\overline{H}} = \frac{H(X)}{\overline{H}}$$

The degree of anonymity thus allows us to quantify the general anonymity of every node in the anonymity network. This is not the only definition, as for instance [DSD04] provides an unbounded definition compared to our bounded definition from [SD03].

2.2 Probable Innocence

While the degree of anonymity is a global quantity assigned to an anonymity network, *probable innocence* is a concept applied to individual members of a network. Deciding if a node is the initiator or not can be interpreted as flipping an unfair coin. If the result is skewed toward one side, the decision if a given node is probably guilty or probably innocent is possible to make.

Definition 2.2 (probable innocence)**.** *A node n_i is probably innocent (of sending the message) if and only if $p_i \leq \frac{1}{2}$.*

A probably innocent node is especially able to deny any communication originated from it. Probable innocence is thus synonymous to deniable communication.

2.3 Message Passing

The method of message passing has not been discussed yet and only established that it is based on a statistical method. Given the strict boundaries we have established for our anonymity network, especially the assumption that corrupt nodes are not distinguishable, allows just one canonical way of handling messages in each node. We can not prefer one over another node and thus have to handle messages uniformly no matter who it came from. This directly leads us to the 'Crowds' anonymity network in the next section.

3 Crowds

The 'Crowds' anonymity network was proposed in [RR98] and handles message passing using a *forwarding probabilty* $\lambda \in [0, 1]$, which is the only parameter of this anonymity network.

If a message is received, the node flips a biased coin with $\mathbb{P}(\text{Heads}) = \lambda$. If it yields 'Heads', it forwards the message to a uniformly chosen node. If the coin yields 'Tails', the node sends the message to the specified receiver.

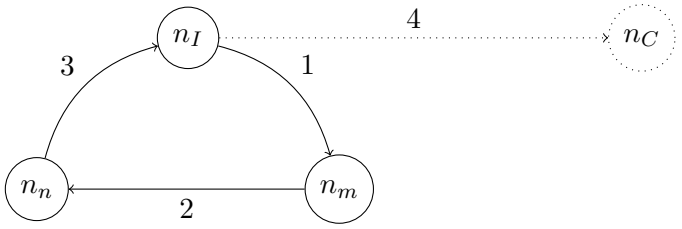

Figure 3.1: Example of a path where the real message initiator n_I occurs twice and immediately precedes the first corrupt node n_C, which is a counterexample for $I \to H_1$.

3.1 Events

We introduce a few notations for events based on the path a message takes in the anonymity network up to the point we intercepted it. Let $k > 0$.

$$\begin{aligned} H_k &:= \text{first corrupt node is at the } k\text{th path-position} \\ H_{k+} &:= \bigvee_{i=k}^{\infty} H_i \\ I &:= \text{first corrupt node immediately postcedes the real message initiator} \end{aligned}$$

It follows trivially that $H_1 \to I$, but $I \not\to H_1$, as the real message initiator can appear on the path multiple times (see figure 3.1).

3.2 Probable Innocence

As we can only examine the node that forwarded the message to one of our corrupt nodes, we define probable innocence if and only if one of our corrupt nodes is on the path (event H_{1+}) and the real message initiator can deny being the initiator even if it had become the suspect message initiator by directly sending the message to a corrupt node (event I).

Definition 3.1 (probable Crowds-innocence)**.** *The suspect message initiator is probably Crowds-innocent if and only if*

$$\mathbb{P}(I|H_{1+}) \leq \frac{1}{2}.$$

Proposition 3.2.

$$\mathbb{P}(I|H_{1+}) = \frac{N - \lambda \cdot (N - C - 1)}{N}$$

Proof. For the first corrupt node to be at the kth path-position, the message first has to be forwarded to $k-1$ good nodes and then forwarded to the corrupt node itself.

$$\mathbb{P}(H_k) = \left(\lambda \cdot \frac{N-C}{N}\right)^{k-1} \cdot \left(\lambda \cdot \frac{C}{N}\right) = \lambda^k \cdot \left(\frac{N-C}{N}\right)^{k-1} \cdot \frac{C}{N}$$

It follows for the compound probability

$$\begin{aligned}
\mathbb{P}(H_{k+}) &= \sum_{i=k}^{\infty} \mathbb{P}(H_i) \\
&= \sum_{i=k}^{\infty} \lambda^i \cdot \left(\frac{N-C}{N}\right)^{i-1} \cdot \frac{C}{N} \\
&= \frac{C}{N} \cdot \frac{N}{N-C} \cdot \sum_{i=k}^{\infty} \left(\lambda \cdot \frac{N-C}{N}\right)^i \\
&= \frac{C}{N-C} \cdot \sum_{i=k}^{\infty} \left(\lambda \cdot \frac{N-C}{N}\right)^i.
\end{aligned}$$

We define $q := \lambda \cdot \frac{N-C}{N}$ and note that $q \in [0,1)$, allowing us to apply the finite and infinite geometric series after splitting up the term.

$$\begin{aligned}
&= \frac{C}{N-C} \cdot \left(\sum_{i=0}^{\infty} q^i - \sum_{i=0}^{k-1} q^i\right) \\
&= \frac{C}{N-C} \cdot \left(\frac{1}{1-q} - \frac{1-q^{(k-1)+1}}{1-q}\right) \\
&= \frac{C}{N-C} \cdot \left(\frac{q^k}{1-q}\right) \\
&= \frac{C \cdot \left(\lambda \cdot \frac{N-C}{N}\right)^k}{(N-C) \cdot \left(1 - \lambda \cdot \frac{N-C}{N}\right)}
\end{aligned}$$

We also know that $\mathbb{P}(I|H_1) = 1$, as $H_1 \to I$, and $\mathbb{P}(I|H_{2+}) = \frac{1}{N-C}$, which is the probability that the the first corrupt node got its message forwarded to by the real message initiator later in the path, which is one among the $N - C$ good nodes (see figure 3.1). Using the law of total probabilty, we determine $\mathbb{P}(I)$.

$$\begin{aligned}
\mathbb{P}(I) &= \mathbb{P}(H_1) \cdot \mathbb{P}(I|H_1) + \mathbb{P}(H_{2+}) \cdot \mathbb{P}(I|H_{2+}) \\
&= \left[\lambda \cdot \frac{C}{N}\right] \cdot 1 + \left[\frac{C \cdot \left(\lambda \cdot \frac{N-C}{N}\right)^2}{(N-C) \cdot \left(1 - \lambda \cdot \frac{N-C}{N}\right)}\right] \cdot \left[\frac{1}{N-C}\right] \\
&= \lambda \cdot \frac{C}{N} + \frac{C \cdot \lambda^2}{N^2 \cdot \left(1 - \lambda \cdot \frac{N-C}{N}\right)} \\
&= \frac{\lambda \cdot C}{N} \cdot \left(1 + \frac{\lambda}{N - \lambda \cdot (N-C)}\right)
\end{aligned}$$

Consequently, using the definition of conditional probability, we determine $\mathbb{P}(I|H_{1+})$.

$$\mathbb{P}(I|H_{1+}) = \frac{\mathbb{P}(I \wedge H_{1+})}{\mathbb{P}(H_{1+})}$$

By construction we know trivially, that $I \to H_{1+}$, as knowing that the first corrupt node immediately postcedes the real message initiator implies that one corrupt node is on the path, and thus $\mathbb{P}(I \wedge H_{1+}) = \mathbb{P}(I)$ hold.

$$
\begin{aligned}
&= \frac{\mathbb{P}(I)}{\mathbb{P}(H_{1+})} \\
&= \frac{\lambda \cdot C}{N} \cdot \left(1 + \frac{\lambda}{N - \lambda \cdot (N - C)}\right) \cdot \frac{(N - C) \cdot \left(1 - \lambda \cdot \frac{N-C}{N}\right)}{C \cdot \left(\lambda \cdot \frac{N-C}{N}\right)^1} \\
&= \frac{\lambda \cdot C}{N} \cdot \left(1 + \frac{\lambda}{N - \lambda \cdot (N - C)}\right) \cdot \frac{(N - C) \cdot N}{C \cdot \lambda \cdot (N - C)} \cdot \left(1 - \lambda \cdot \frac{N - C}{N}\right) \\
&= \left(1 + \frac{\lambda}{N - \lambda \cdot (N - C)}\right) \cdot \left(1 - \lambda \cdot \frac{N - C}{N}\right) \\
&= 1 - \lambda \cdot \frac{N - C}{N} + \frac{\lambda}{N - \lambda \cdot (N - C)} \cdot \left(1 - \lambda \cdot \frac{N - C}{N}\right) \\
&= 1 - \lambda \cdot \frac{N - C}{N} + \frac{\lambda}{N - \lambda \cdot (N - C)} \cdot \frac{1}{N} \cdot (N - \lambda \cdot (N - C)) \\
&= 1 - \lambda \cdot \frac{N - C}{N} + \frac{\lambda}{N} \\
&= \frac{N - \lambda \cdot (N - C - 1)}{N}
\end{aligned}
$$

□

Corollary 3.3. *The suspect message initiator is probably Crowds-innocent if and only if* $\lambda > \frac{1}{2}$ *and*

$$N \geq \frac{1}{1 - \frac{1}{2 \cdot \lambda}} \cdot (C + 1).$$

Proof.

$$
\begin{aligned}
\frac{1}{2} \geq \mathbb{P}(I|H_{1+}) &\overset{3.2}{=} \frac{N - \lambda \cdot (N - C - 1)}{N} \\
\Leftrightarrow \frac{N}{2} - N + \lambda \cdot N &\geq -\lambda \cdot (-C - 1) \\
\Leftrightarrow N \cdot \left(\lambda - \frac{1}{2}\right) &\geq \lambda \cdot (C + 1)
\end{aligned}
$$

Now we are using the fact that we assume $\lambda > \frac{1}{2}$, which implies that $(\lambda - \frac{1}{2}) > 0$.

$$
\begin{aligned}
\Leftrightarrow N &\geq \frac{\lambda}{\lambda - \frac{1}{2}} \cdot (C + 1) \\
\Leftrightarrow N &\geq \frac{1}{1 - \frac{1}{2 \cdot \lambda}} \cdot (C + 1)
\end{aligned}
$$

□

Corollary 3.4. *The* $N - C - 1$ *good nodes besides the suspect message initiator have probability*

$$\frac{\lambda}{N}$$

of being the message initiator and are probably Crowds-innocent.

Proof. The remaining $N - C - 1$ nodes, assuming we don't have any additional information about the network, distribute the remaining probability $1 - \mathbb{P}(I|H_{1+})$ uniformly among them. We obtain for them that

$$\frac{1 - \mathbb{P}(I|H_{1+})}{N - C - 1} = \frac{N - N + \lambda \cdot (N - C - 1)}{N \cdot (N - C - 1)} = \frac{\lambda}{N} < \frac{1}{N} \leq \frac{1}{2},$$

as they only occur when $N \geq 2$, which makes them probably Crowds-innocent in all cases. □

As a final remark, we can see that we want to choose λ as large as possible to make the lower bound for N as large as possible, keeping in mind though that with increasing λ, messages on average have to take increasingly longer paths until they reach their destination, putting more load on the anonymity network in total.

3.3 Degree of Anonymity

Now that we have determined all node probabilities, it is time to determine the degree of anonymity of the 'Crowds' anonymity network.

Proposition 3.5.

$$d(X) = \frac{(N - \lambda \cdot (N - C - 1)) \cdot \ln\left(\frac{N}{N - \lambda \cdot (N - C - 1)}\right) + \lambda \cdot (N - C - 1) \cdot \ln\left(\frac{N}{\lambda}\right)}{N \cdot \ln(N - C)}$$

Proof. To determine the degree of anonymity, we need to assign probabilities to each node in our network. We already established that our C corrupt nodes have probabilities $p_i = 0$. Additionally, our suspect message initiator who sent a message to our first corrupt node in the path has probability $\mathbb{P}(I|H_{1+})$ (see proposition 3.2) and the remaining $N - C - 1$ good nodes have probability $\frac{\lambda}{N}$ (see corollary 3.4) of being the real message initiator. It follows directly for the random variable X as previously established by definition, that

$$\begin{aligned}
H(X) &= -\sum_{i=1}^{N} p_i \cdot \ln(pi) \\
&= -C \cdot [0 \cdot \ln(0)] - 1 \cdot [\mathbb{P}(I|H_{1+}) \cdot \ln(\mathbb{P}(I|H_{1+}))] - \\
&\quad (N - C - 1) \cdot \frac{1 - \mathbb{P}(I|H_{1+})}{N - C - 1} \cdot \ln\left(\frac{1 - \mathbb{P}(I|H_{1+})}{N - C - 1}\right) \\
&= -\frac{N - \lambda \cdot (N - C - 1)}{N} \cdot \ln\left(\frac{N - \lambda \cdot (N - C - 1)}{N}\right) - \frac{\lambda \cdot (N - C - 1)}{N} \cdot \ln\left(\frac{\lambda}{N}\right) \\
&= \frac{N - \lambda \cdot (N - C - 1)}{N} \cdot \ln\left(\frac{N}{N - \lambda \cdot (N - C - 1)}\right) + \frac{\lambda \cdot (N - C - 1)}{N} \cdot \ln\left(\frac{N}{\lambda}\right),
\end{aligned}$$

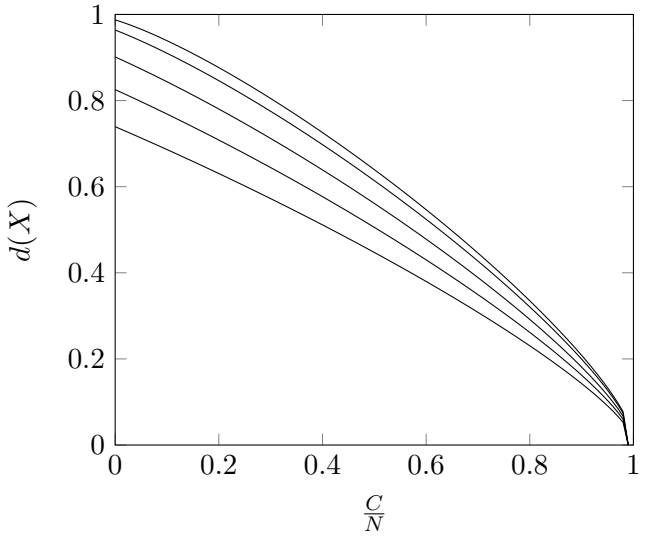

Figure 3.2: $d(X)$ for $\lambda \in \{0.6, 0.7, 0.8, 0.9, 0.95\}$ (bottom to top graphs) over $\frac{C}{N}$.

and by definition for the degree of anonymity

$$\begin{aligned} d(X) &= \frac{H(X)}{\overline{H}} \\ &= \frac{(N - \lambda \cdot (N - C - 1)) \cdot \ln\left(\frac{N}{N - \lambda \cdot (N - C - 1)}\right) + \lambda \cdot (N - C - 1) \cdot \ln\left(\frac{N}{\lambda}\right)}{N \cdot \ln(N - C)}. \end{aligned}$$ □

Given the complexity of the formula for $d(X)$, it makes sense to plot it for different values of λ over $\frac{C}{N}$, a ratio expressing the corruptness of the network (see figure 3.2). As expected, with increasing corruptness, the degree of anonymity decreases to zero.

4 Conclusion

Using statistical methods to obscure the initiator of a message proves to be a viable direction. Even though it has been shown that the Crowds anonymity network is vulnerable to the predecessor attack (see [WALS04]), it is still useful to explain the idea behind anonymity networks, and the analytical methods used to investigate it can be a foundation for the investigation of more elaborate ideas.

In the arms race between those who silence and those who speak, instead of creating dedicated infrastructure for confidential message transmission, it might be the best choice to just blend in with the crowd. Convincing somebody who 'has nothing to hide' of moving his communications to an anonymity network is easier with the knowledge that it helps protect those who depend on it.

References

[DS05] George Danezis and Andrei Serjantov. Statistical disclosure or intersection attacks on anonymity systems. In Jessica Fridrich, editor, *Information Hiding: 6th International Workshop*, pages 293–308. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005. ISBN 978-3-540-30114-1. URL `http://dx.doi.org/10.1007/978-3-540-30114-1_21`.

[DSCP03] Claudia Díaz, Stefaan Seys, Joris Claessens, and Bart Preneel. Towards measuring anonymity. In *Proceedings of the 2nd International Conference on Privacy Enhancing Technologies*, PET'02, pages 54–68. Springer-Verlag, Berlin, Heidelberg, 2003. ISBN 3-540-00565-X. URL `http://dl.acm.org/citation.cfm?id=1765299.1765304`.

[DSD04] Claudia Díaz, Len Sassaman, and Evelyne Dewitte. Comparison between two practical mix designs. In Pierangela Samarati, Peter Ryan, Dieter Gollmann, and Refik Molva, editors, *Computer Security – ESORICS 2004: 9th European Symposium on Research in Computer Security*, pages 141–159. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004. ISBN 978-3-540-30108-0. URL `http://dx.doi.org/10.1007/978-3-540-30108-0_9`.

[RR98] Michael K. Reiter and Aviel D. Rubin. Crowds: Anonymity for web transactions. *ACM Transactions on Information and System Security (TISSEC)*, 1(1):66–92, November 1998. ISSN 1094-9224. URL `http://doi.acm.org/10.1145/290163.290168`.

[SD03] Andrei Serjantov and George Danezis. Towards an information theoretic metric for anonymity. In Roger Dingledine and Paul Syverson, editors, *Privacy Enhancing Technologies: Second International Workshop, PET 2002*, pages 41–53. Springer Berlin Heidelberg, Berlin, Heidelberg, 2003. ISBN 978-3-540-36467-2. URL `http://dx.doi.org/10.1007/3-540-36467-6_4`.

[WALS04] Matthew K. Wright, Micah Adler, Brian Neil Levine, and Clay Shields. The predecessor attack: An analysis of a threat to anonymous communications systems. *ACM Transactions on Information and System Security (TISSEC)*, 7(4):489–522, November 2004. ISSN 1094-9224. URL `http://doi.acm.org/10.1145/1042031.1042032`.