Notes for privacy

Gabriele Genovese

16 December 2024

1 How to share a secret

11 scientists are working on a secret project. They don't trust each other. The project is in a digital safe. To open the digital safe, at least 6 out of the 11 scientists must be present.

Let D be some secret data. Lets divide D into n pieces $D_1, ..., D_n$ such that:

- Knowledge on any k or more D_i pieces makes D easily computable
- Knowledge of any $k-1$ or fewer D_i pieces leaves D completely undetermined

Trivial solution is impratical (too many fragments to manage and time to reconstruct the secret). Let's use polynomial interpolation: a polynomial of degree $k-1$ is uniquely defined by k points; with $k-1$ points only there is an infinity of k polynomial that can cross those points. So you need at least k points to find the polynomial equation $g(x)$ using *Lagrange interpolation*. The secret is g(0).

1.1 Create the n fragments

Let D be your secret (D is an integer without loss of generality). Choose a prime $p > \max(D, n)$. $g(x)$ is a random polynomial of degree $k-1$ so that $g(x) = \sum_{i=0}^{k-1} a_i x^i$, where $a_0 = D$ and $a_i,\ i \in \{1, ..., k-1\}$ are chosen with a uniform distribution on $[0, p[$.

$$g(x) = a_0 + a_1 x + a_2 x^2 + ... + a_{k-1} x^{k-1}$$

Compute: $D_1 = g(1) \bmod p$, $D_2 = g(2) \bmod p$, ..., $D_n = g(n) \bmod p$. Distribute the tuples (i, D_i).

1.2 Retrive D based on k fragments (x_i, D_i)

Use Lagrange polynomial interpolation to reconstruct the polynomial. $g(0) = D$ is the secret.

$$g(0) = \sum_{i=1}^{k} D_i \left(\prod_{j=1, j\neq i}^{k} \frac{-x_j}{x_i - x_j} \right)$$

1.3 Example for $(k = 3, n = 5)$

The secret is $D = 148$. Let's take $p = 997$ (prime), $a_1 = 59$ and $a_2 = 340$ (random) such that $g(x) = 148 + 59x + 340x^2$. We compute 5 fragments:

- $D_1 = g(1) \bmod 997 = 547$
- $D_2 = g(2) \bmod 997 = 1626 \bmod 997 = 629$
- $D_3 = g(3) \bmod 997 = 3385 \bmod 997 = 394$
- $D_4 = g(4) \bmod 997 = 5824 \bmod 997 = 839$
- $D_5 = g(5) \bmod 997 = 8943 \bmod 997 = 967$

We give to each user a fragment among $(1, 547), (2, 629), (3, 394), (4, 839), (5, 967)$.

Assume users with fragments 1,3,4 want to reconstruct the secret:

$$g(0) = 547 \cdot \left(\frac{-3}{1-3} \cdot \frac{-4}{1-4}\right) + 394 \cdot \left(\frac{-1}{3-1} \cdot \frac{-4}{3-4}\right) + 839 \cdot \left(\frac{-1}{4-1} \cdot \frac{-3}{4-3}\right) =$$

$$= 547 \cdot \left(\frac{3}{2} \cdot \frac{4}{3}\right) + 394 \cdot \left(-\frac{1}{2} \cdot 4\right) + 839 \cdot \left(-\frac{1}{3} \cdot -3\right) =$$

$$= 547 \cdot 2 + 394 \cdot -2 + 839 = 1145 \bmod 997 = 148$$

1.4 Properties

- The size of each fragment does not exceed the size of the secret (if p is the same size order as the secret)
- New fragments can be generated at any time without affecting existing ones
- All fragments can be changed without changing the secret by generating a new polynomial
- Possibility of hierarchical schemes by giving a different number of fragments depending on roles (e.g., president 3 fragments, executives 1 fragment)
- No unproven assumptions (unlike cryptographic or hash protocols)

2 Chaum-net and Mix-net

It's the basis for onion routing. Alice wants to send a message M to Bob. Assume an unsecure communication network. Nobody knows who is the sender (even Bob). Nobody knows who is the receiver (except Alice). Nobody, except Bob, is able to get M. Use RSA cryptographic system ($K(K^{-1}(M)) = K^{-1}(K(M)) = M$ where K is public).

2.1 Sealed message

Create a large random string R (e.g., 256 bits large). Append R to the message $M : R, M$. The sealed message is $K(R, M)$. As R is a large random string, not practical to guess R, M. Once Bob get $K(R, M)$. Compute $K^{-1}(K(R, M)) = R, M$. Remove R (easy if R is fixed length).

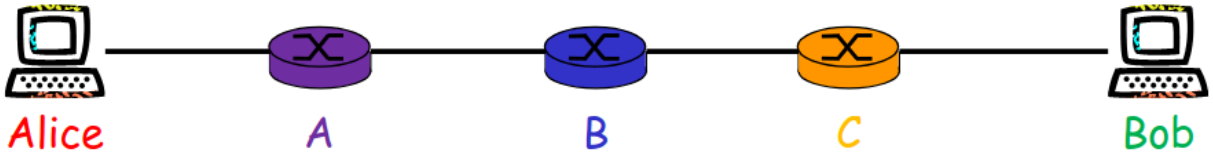
2.2 Mix

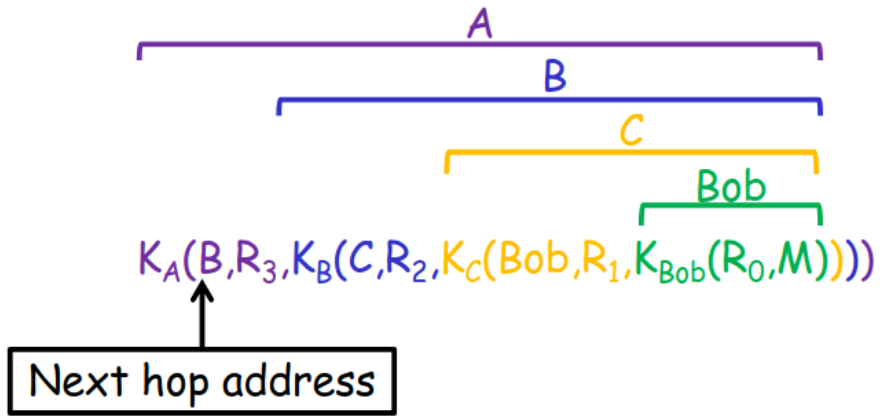
A mix is a machine (dedicated machine, router, end-user in an overlay). Its purpose is to to hide correspondences between incoming and outgoing messages. Not possible to map a source and an outgoing message (apart for the mix). No possible to map a receiver and an incoming message (apart for the mix). If the mix is compromised, it's possible to know the sender and receiver for each message but, it's impossible to find what is the message.

Use a cascade of mixes: it works with a partially trusted set of mixes as long as one mix in the cascade can be trusted or, as long as all untrusted mixes in the cascade do not cooperate.

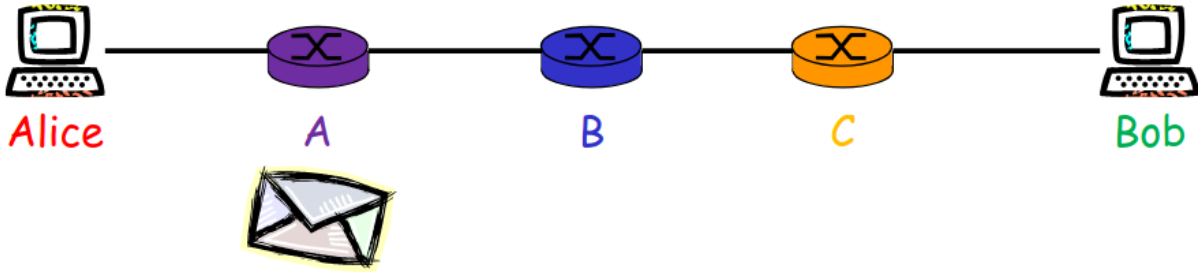
Increasing the number of mixes in the cascade can increase the confidence but, increases the end-to-end delay. Tor uses at least 3 mixes selected at random, called a „*Circuit*“. Periodically select new random mixes to form a new circuit.

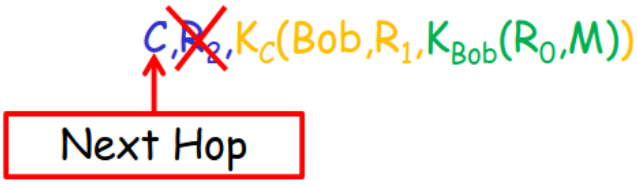
2.3 How it works





50





52

2.4 Attacks

Are still possible:

- timing
- active end-users

3 Privacy

You should care about it because the concept of morality is different in every country.

Activity is linked to **identity**.

Anonymization is breaking this link.

3.1 Identity

Any network layer can be considered activity (Web history, IP, MAC, BitTorrent download history, VoIP conversation).

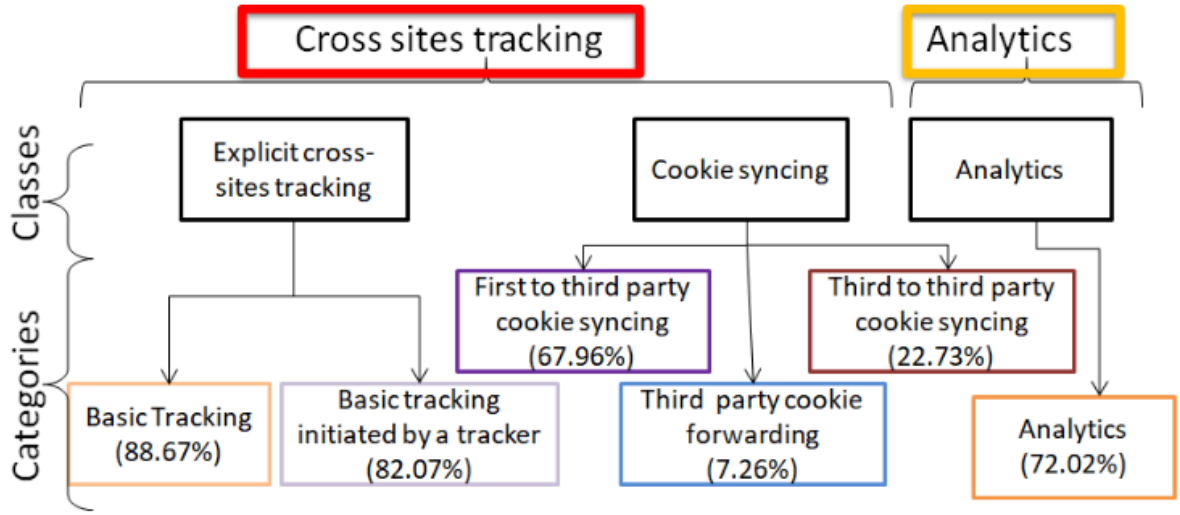
Three different notion of identities:

- *Network* identity: An IP address in IP networks; Enables network profiling
- *Application* identity: Temporary (cookie) or permanent (skype ID)
- *Social* identity: Everything that identify a user in real life (Name, postal address, email)

3.2 Suoi papers

4 Traking

Tracking can be exploited to targeted advertisement and manipulation (biggest case Cambrige Analytica).



Google analytics is one of the most used analytics tracking tool.

Invisible pixels.