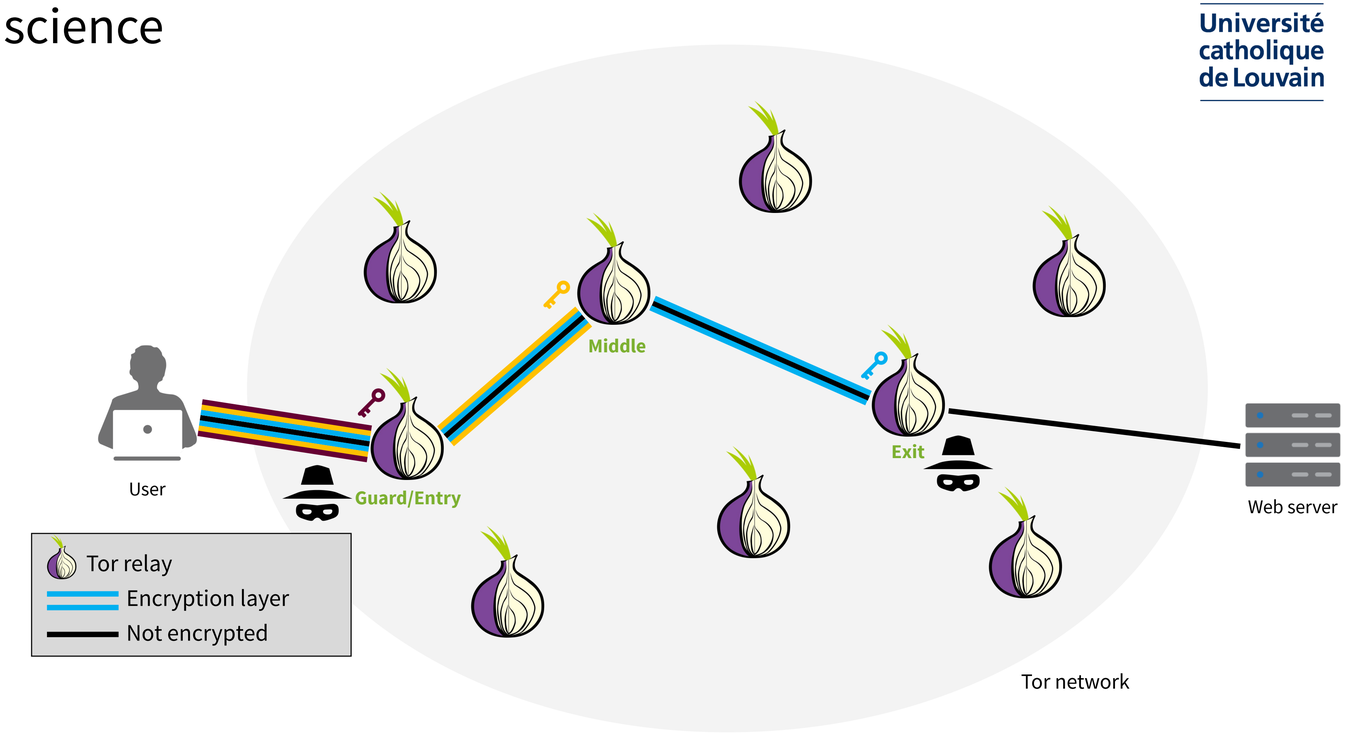
Improving anonymity in Tor through diversity

Robin Descamps
Université Catholique de Louvain — Computer science
Supervisor: Olivier Pereira

What is Tor?

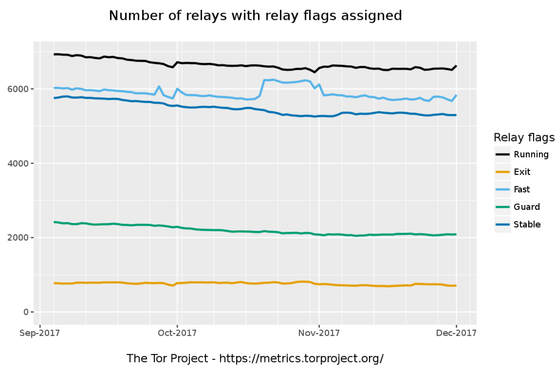
Tor is a low-latency overlay network in which users connect to TCP services through a series of nodes (an **entry relay**, a **middle relay**, and an **exit relay**) which allows them to gain anonymity. These nodes, called relays, are running by volunteers accross the world. Anyone can set up a Tor relay, and define constraints on it as he wishes (e.g. a relay that cannot be an exit).

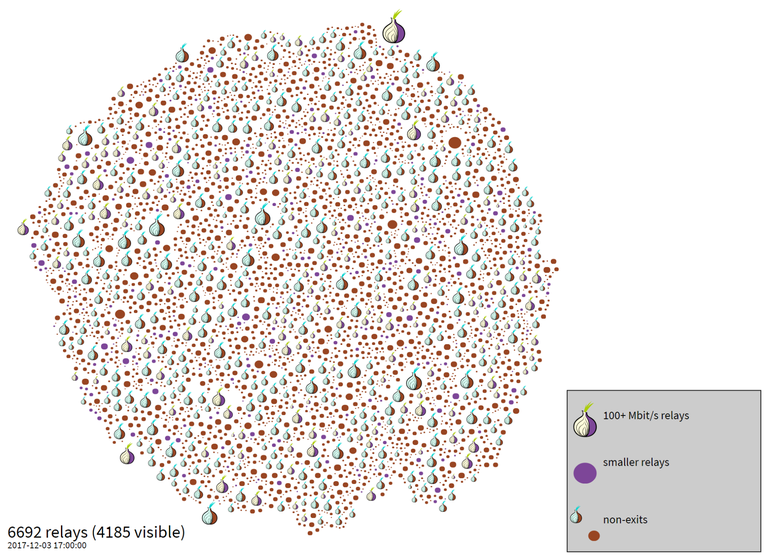
Tor is by design vulnerable to **end-to-end traffic correlation attacks**, i.e. matching traffic at both ends of a circuit. Thus, an adversary that can observe traffic entering and leaving the Tor network is able to de-anonymize users with a certain probability over a certain period of time, depending of its ressources (number of nodes/bandwidth under control in the Tor network).



Objectives

This master thesis aims to **define methods to quantify the diversity of the nodes** in the Tor network. Increasing the diversity will increase the anonymity of the Tor users, and will mitigate traffic correlation attacks. Existing metrics do not allow us to quantify the diversity of the nodes (see figures at the right).





Methodology

1. Define the adversaries

Relay adversary: Control of a part of the Tor network, by adding or corrupting nodes.

Network adversary: Control of an AS, IXP, Cloud service provider, or some other authority

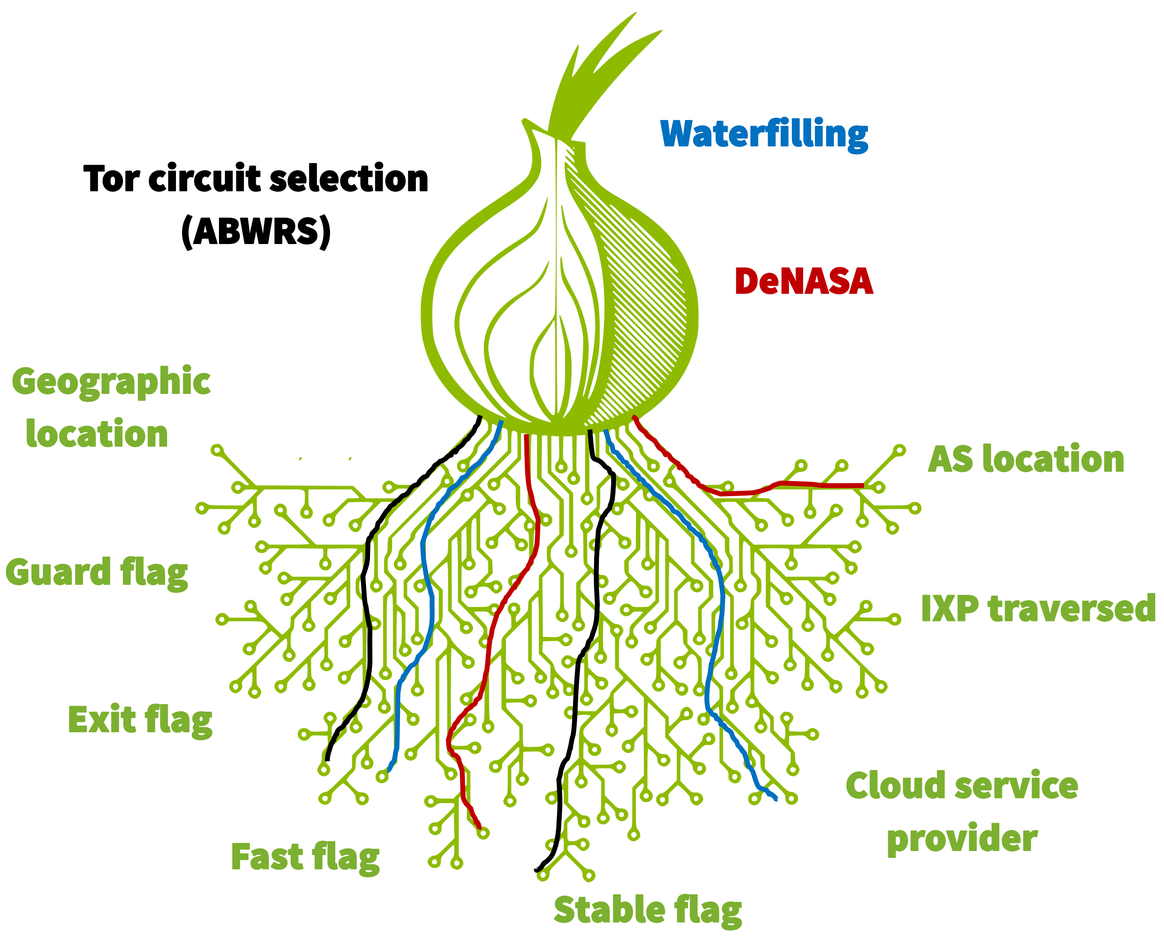
2. Define security metrics

According to defined adversaries

- **Network entropy**: The *guessing entropy*;
- Numbers derivated from the probability distribution on **number of paths compromises over a certain period of time**;
- Numbers derivated from the probability distribution on **time before the first path compromise**.

3. Simulate the network

- Under **different path selection algorithms**
- Before and after adding groups of nodes with **different parameters**



4. Analyze the results

Which parameter has the most impact on anonymity? Which parameters configuration should we encourage the new volunteers to use? Under which path selection algorithm do we achieve the most diversity?

Furthermore, is it possible and/or costly to deploy such a node? (for example, set up a relay in China is not possible). We must also take the volunteer constraints into account, so we would be able at last to **propose** him **a configuration that contribute the most to the network diversity, under budget and user constraints**. A rewarding system could also as well be deployed in such a context.

Tools

- **CollecTor** : Gather network state archives
- **TorPS** : Simulate Tor network against adversary with selected path selection
- **Shadow** : Simulate Tor network under real conditions

Planning

- Read and analyze all previous work related to anonymity improvement in the Tor network, with a focus on the diversity of its relays.
- Adapt and optimize the TorPS tool to integrate all the parameters to change in the simulations to perform.
- Perform as much as possible simulations with TorPS (and then possibly with Shadow) with the different path selections algorithms and relays parameters.
- Analyze the results so that we are able to define a « best » relay configuration to add to the Tor network, under the existing network, path selection algorithm, and economic/political context and user budget/constraints.