Improving anonymity in Tor through diversity
Master thesis plan

Descamps Robin

November 17, 2017

1 Context

This thesis aims to contribute to the Tor project. Tor is a low-latency anonymity 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).

Nowadays, the Tor network performs a good trade-off between performance and anonymity provided, thanks to the selection path algorithm. However, some attacks allow adversaries to de-anonymize users with a non-negligable chance over a reasonable period of time by observing and matching traffic at both ends of the user communication. The purpose of this thesis is to contribute to the Tor project by proposing methods in order to improve the anonymity of their users, and thus to mitigate these attacks.

2 Objectives

The objective of this thesis is to define several methods to quantify the diversity of the nodes deployed in the Tor network. We can then use this model to define its security level, and evaluate the utility of new relays, depending on their characteristics: location, bandwidth, Tor flags,...

This thesis could thus have several impacts. First, contribute to the Tor metrics by proposing new metrics about anonymity level we can obtain by using the network, thanks to the diversity of its nodes. Secondly, we are able to attribute an utility to each relay in the Tor network, depending on its contribution to the network diversity, so that a rewarding system could be set up.

3 Methodology

In order to create a method to measure the diversity of the relays in the Tor network, we will define metrics to measure this diversity, with respect to certain adversaries. We first define the two considered adversaries in this context [15]:

- Relay adversary: An adversary who possess a "bandwidth power", either by corrupting existing relays, or by contributing to the Tor network with relays under his control;
- Network adversary: An adversary who control an Autonomous System, an Internet Exchange Point, a cloud service provider, or even a country or some other authority. This kind of adversary can also control several of them, and is here considered able to see all their Tor traffic.

In either case of adversary, we can in a realistic way consider that they have a purpose, e.g. de-anonymize a certain class of users [15].

Then, we define the security metrics we will use:

- Network entropy: The guessing entropy [20]
- The probability distribution on number of path compromises on a certain period of time;
- The probability distribution on time before the first path compromise.

For the two last metrics, the metrics derivated from the distributions must still be defined later. We will analyze the distributions, but in a second time we will need to derive numbers from the constructed distributions in order to compare them.

The next step is to simulate the end-to-end correlation attack under the path selection algorithm of Tor, as well as other path selection algorithm with the TorPS tool [6]. The Shadow tool [2] can also be useful here, as it can simulate network information that we cannot retrieve on Tor usage (such as relay congestion). We first reproduce at best the real Tor network, and define a typical user model to generate streams. We use the Monte Carlo method in order to have a representative set of data that reproduce the real behaviour of the live Tor network.

The different path selection algorithm we will use are:

- The Tor selection circuit (ABWRS) [10];
- The DeNASA selection circuit [7];
- The Waterfilling selection circuit [20].

These path selection methods are the most recent ones to propose an anonymity improvement for the Tor users form the point of view of the adversaries we described above.

We then perform the core of the experimentation: we add relays to the network. We analyze the impact on the metrics described above in function of the following parameters:

- The relay flags: entry, middle, exit, fast, stable;
- The AS location;

- The cloud service provider
- The geographic location.

By changing all these parameters, one at a time, we can compare the results: which of them has most impact on anonymity than others? Which of them has negligible impact? Which of the parameter configuration should we encourage the new volunteers to use?

Furthermore, is it possible to deploy such a relay (for example: set up a relay in China is not possible)? Or is it worth, i.e. create a relay in Africa which may be much more expensive than creating it in Europe. Then, we can derive a utility from chosen parameters of a new relay, but we must take the relay price into account. In addition to that, we must also look at the user constraints in our calculation (e.g. does he want to create an exit relay?).

According to the adversary we want to "counter", we may have several choices. Relay configuration can improve security against relay adversary more than against network adversary, or vice-versa. At this step, we could optionally ponderate the adversaries to give a single score of security according to relay parameters (to give more usability). Therefrom, we sould be able to propose the best relay location and flags according to the user budget and preferences.

4 Resources

First, the thesis lays mainly on the documentation, the specifications and the metrics, data,... provided by the Tor project website [3], and especially from the CollecTor [1] platform. For the technical aspects, the TorPS [6] and the Shadow simulator [2] tools are useful to conduct the experiments described above. The active Tor community will also be consulted, particularly the metrics and the dev teams. Finally, the recent literature about Tor or anonymous systems in general is also essential to set a state of the art of the diversity aspect about the Tor network. A first bibligraphy is provided below.

5 Planning

Here is a proposal of an inital planning of all the tasks to complete:

Date	Task
Before December	Production of a first version of the poster presented the 4th december to MA1 students, researchers and professors at UCL
End of December	First simulations with preliminary results about diversity level
End of February	Simulations complete, with full results about diversity obtained through all defined parameters
Before Easter Holiday	Results, analysis and relay rewarding system proposition
End of April	Thesis full draft, a first version
Mid-May	Thesis full draft, an improved version
End of May	Thesis final version
Before June	Presentation slides

References

[1] CollecTor. https://collector.torproject.org/, 2017.

[2] The Shadow simulator. https://shadow.github.io/, 2017.

[3] The Tor Project. https://torproject.org/, 2017.

[4] Tor directory protocol, version 3. https://gitweb.torproject.org/torspec.git/tree/dir-spec.txt, 2017.

[5] Tor Metrics Portal. https://metrics.torproject.org/, 2017.

[6] TorPS: The Tor path simulator. https://torps.github.io/, 2017.

[7] Armon Barton and Matthew Wright. Denasa: Destination-naive as-awareness in anonymous communications. *Proceedings on Privacy Enhancing Technologies*, 2016(4):356–372, 2016.

[8] Nikita Borisov, George Danezis, Prateek Mittal, and Parisa Tabriz. Denial of service or denial of security? In *Proceedings of the 14th ACM conference on Computer and communications security*, pages 92–102. ACM, 2007.

[9] Roger Dingledine and Nick Mathewson. Tor protocol specifications. https://gitweb.torproject.org/torspec.git/tree/tor-spec.txt, 2017.

[10] Roger Dingledine, Nick Mathewson, Steven Murdoch, and Paul Syverson. Tor: The second-generation onion router (2014 draft v1), 2014.

[11] Tariq Elahi, Kevin Bauer, Mashael AlSabah, Roger Dingledine, and Ian Goldberg. Changing of the guards: A framework for understanding and improving entry guard selection in tor. In *Proceedings of the 2012 ACM Workshop on Privacy in the Electronic Society*, pages 43–54. ACM, 2012.

[12] Nick Feamster and Roger Dingledine. Location diversity in anonymity networks. In *Proceedings of the 2004 ACM workshop on Privacy in the electronic society*, pages 66–76. ACM, 2004.

[13] Angele Hamel, Jean-Charles Grégoire, and Ian Goldberg. The misentropists: New approaches to measures in tor. *Centre for Applied Cryptographic Research (CACR)*, 2011.

[14] Nicholas Hopper, Eugene Y Vasserman, and Eric Chan-Tin. How much anonymity does network latency leak? *ACM Transactions on Information and System Security (TISSEC)*, 13(2):13, 2010.

[15] Aaron Johnson, Chris Wacek, Rob Jansen, Micah Sherr, and Paul Syverson. Users get routed: Traffic correlation on tor by realistic adversaries. In *Proceedings of the 20th ACM Conference on Computer and Communications Security (CCS 2013)*. ACM, 2013.

[16] Aaron M Johnson, Paul Syverson, Roger Dingledine, and Nick Mathewson. Trust-based anonymous communication: Adversary models and routing algorithms. In *Proceedings of the 18th ACM conference on Computer and communications security*, pages 175–186. ACM, 2011.

[17] Joshua Juen, Aaron Johnson, Anupam Das, Nikita Borisov, and Matthew Caesar. Defending tor from network adversaries: A case study of network path prediction. *Proceedings on Privacy Enhancing Technologies*, 2015(2):171–187, 2015.

[18] Steven J Murdoch and Robert NM Watson. Metrics for security and performance in low-latency anonymity systems. In *International Symposium on Privacy Enhancing Technologies Symposium*, pages 115–132. Springer, 2008.

[19] Lasse Overlier and Paul Syverson. Locating hidden servers. In *Security and Privacy, 2006 IEEE Symposium on*, pages 15–pp. IEEE, 2006.

[20] Florentin Rochet and Olivier Pereira. Waterfilling: Balancing the tor network with maximum diversity. *Proceedings on Privacy Enhancing Technologies*, 2017(2):4–22, 2017.

[21] Fatemeh Shirazi, Matthias Goehring, and Claudia Diaz. Tor experimentation tools. In *Security and Privacy Workshops (SPW), 2015 IEEE*, pages 206–213. IEEE, 2015.

[22] Yixin Sun, Anne Edmundson, Laurent Vanbever, Oscar Li, Jennifer Rexford, Mung Chiang, and Prateek Mittal. Raptor: Routing attacks on privacy in tor. In *USENIX Security Symposium*, pages 271–286, 2015.

[23] Paul Syverson. Why i'm not an entropist. In *International Workshop on Security Protocols*, pages 213–230. Springer, 2009.

[24] Paul Syverson, Gene Tsudik, Michael Reed, and Carl Landwehr. Towards an analysis of onion routing security. In *Designing Privacy Enhancing Technologies*, pages 96–114. Springer, 2001.

[25] Chris Wacek, Henry Tan, Kevin S Bauer, and Micah Sherr. An empirical evaluation of relay selection in tor. In *NDSS*, volume 13, pages 24–27, 2013.