Statistical Analysis and Anonymity of TOR's Path Selection

Andrei Mogage[1] and Emil Simion[2]

[1]Department of Computer Science, UAIC IASI: mogage.andrei.catalin@info.uaic.ro
[2]University Politehnica of Bucharest, emil.simion@upb.ro

Abstract. Tor [1] is a network based on the onion routing infrastructure [2] and provides many advantages, including tracking avoidance, research, wider access and, unfortunately, illegal activities. To achieve this, the client will connect to a TOR circuit consisting of nodes chosen under certain restrictions. The purpose of this paper is to draw attention of the narrow range of available and constraints obedient nodes. This is of interest because it impacts the anonymity and the privacy of users and their internet traffic.

Keywords: tor · onion routing· cryptography · cyber security

1 Introduction

Tor Network helps users maintain not only their security, but their anonymity and integrity as well. To achieve this, users will construct a circuit consisting of other nodes. By nodes, we refer here to tor routers, meaning relays and bridges. The last node, known as Exit Node [3], will be the one sending the original message to the final destination, i.e. a Web Service. Based on Onion Routing, the user's message will be encrypted, with other node's public key sequentially, each of them decrypting a layer, thus the "onion" notation. Figure 1 represents a short schematic of a packet transfer. *T1, T2, T3* represent the public key of each node, used to incrementally encrypt the message, *M*.

2 Path Selection

Authors at Tor Project provide numerous documents on the way tor works and transfers data. One particular specification which is of interest is path selection and its constraints. Despite being initially mentioned as "Randomly Chosen", the path is determined based on rules. According to the Tor Specifications [4], the exit node is chosen first. This is because the availability of a node able to connect to the user's destination is required and no path can be built without it.
There are two main cases which might occur:

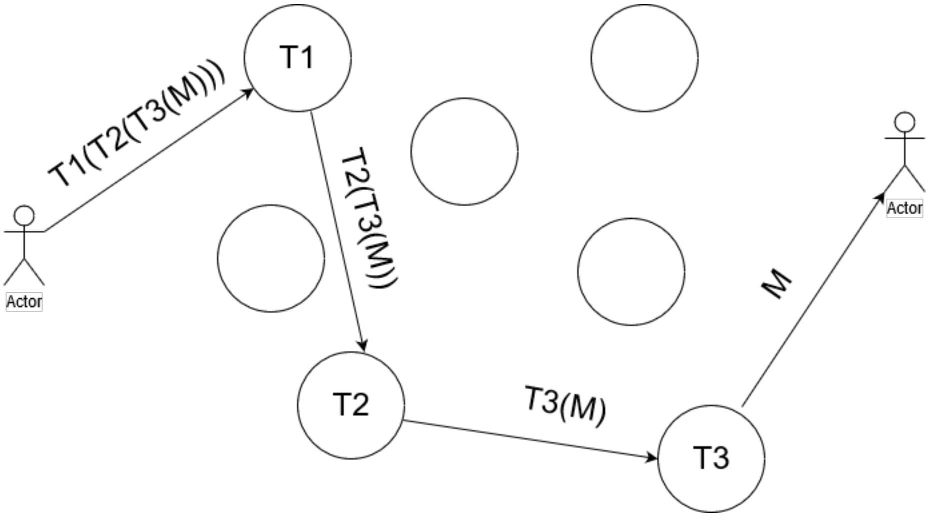

Fig. 1. Onion Routing

- The target IP is known, in which case the exit node will simulate the connection to determine if it will be used as such
- The target IP is unknown, in which case the node will be picked if it ”might support” the connection, based on accepted connections to the specified port

Another rule for the exit node is not to be marked as ”Bad Exit” by other nodes. This happens in cases of nodes used for attacks, tracking, analysis, etc. Apart from that, all nodes which might constitute a path will have to respect certain constraints. This is mainly used as a measure to avoid statistical analysis and tracking:

- The same path cannot use the same nodes.
- No two nodes can be in the same family. This consists of nodes controlled by the same group and, thus, might pose a risk if two nodes are in the same family.
- No two nodes are allowed to be in the same /16 subnet.
- The chosen nodes must be valid and running.
- Guard nodes are those having a ”privileged” position and, thus, the first one must be a Guard Node because it will see the user’s real IP address.
- The length of the path must follow certain restrictions. By default, it is created using 3 nodes.

Besides the constraints, the circuits will be weighted according to ”Computing Bandwidth Weights”[5], in order to increase security and speed. Using Tor Metrics [7], we may provide statistics on actual numbers of node which will be chosen in the creation of a path. Therefore, we have:

- Between 6000 and 7000 running relays and approximate 1000 bridges;

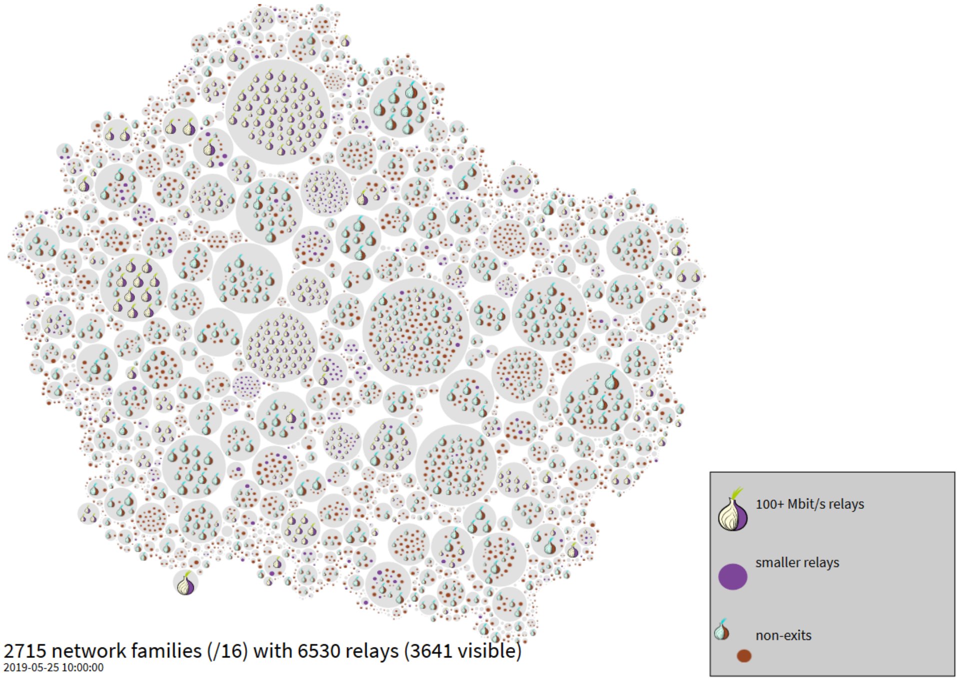

Fig. 2. Network family - all relays [8]

- 1 router per /16 subnet, which leads to max 1 out of 65536 possibilities;
- Less than 1000 exit nodes.

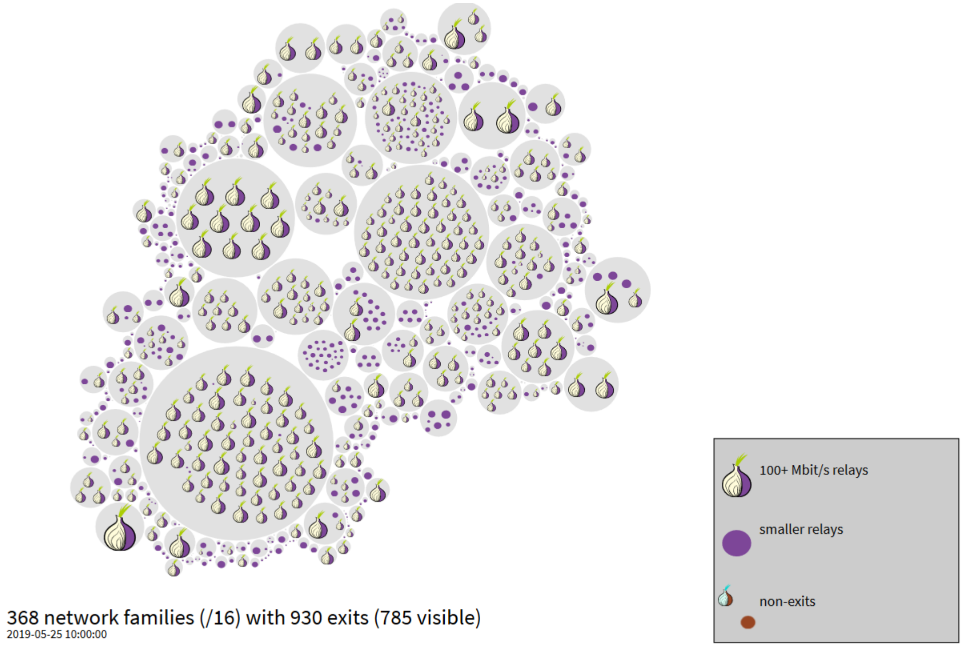

Fig. 3. Network family - exit only [9]

Making a blind assumption of 3 nodes per family and 5 nodes per subnet, leads to a total of 160000 nodes and 1000 exit nodes, meaning 160000000 in total. However, in reality, the number is much smaller. According to [8], there are around 2700 network families (/16) with approximate 6500 relays, leading to a small range of possible nodes. Figures 2 and 3 represent the number of running nodes and their families.

3 Conclusions and future work

Authors at Tor Project provide a research section[10] which invites researchers and others to discuss and provide various ideas. This paper set its main purpose as to draw attention to the low range of possibilities of node selection for path construction. This poses a threat because it increases the chance of statistical based attacks, breaking the anonymity and, in some cases, the entire security of users. What is more, low number of nodes increases latency times and the chance of unsuccessful connections.

Our next steps will focus on providing ways of securing the path construction and, therefore, increasing the security. This will also help in other, but similar, attacks, as website fingerprinting attack.

References

1. https://2019.www.torproject.org/docs/onion-services
2. D. Goldschlag, M. Reed, P. Syverson, "Onion Routing for Anonymous and Private Internet Connections", 1999
3. R.Dingledine, N. Mathewson, P.Syverson, "Tor: The Second-Generation Onion Router"
4. https://gitweb.torproject.org/torspec.git/plain/path-spec.txt
5. https://gitweb.torproject.org/torspec.git/plain/dir-spec.txt
6. https://2019.www.torproject.org/docs/tor-manual.html.en#MyFamily
7. https://metrics.torproject.org
8. https://metrics.torproject.org/bubbles.html#network-family
9. https://metrics.torproject.org/bubbles.html#network-family-exit-only
10. https://research.torproject.org/ideas/