Measuring the safety of the Tor network

Roger Dingledine
arma@torproject.org

Tor Tech Report 2011-02-001
February 6, 2011

0 Introduction

We need a better understanding of how much anonymity the Tor network provides against a partial network adversary who observes and/or operates some of the network. Specifically, we want to understand the chances that this class of adversary can observe a user's traffic coming into the network and also the corresponding traffic exiting the network.

Most of our graphs historically have looked at the total number of relays over time. But this simple scalar doesn't take into account relay capacity, exit policies, geographic or network location, etc.

One concrete problem from this bad metric is that when many new relays show up (e.g. due to a political event[1]), we really don't have any insight about how the diversity of the new relays compares to the rest of the network. The simplistic metric also gets misused in academic research papers to produce misleading results like "I can sign up 2% of the Tor relays and capture 50% of the paths," when what they mean is "I can provide half the capacity in the Tor network and capture half the paths."

There are four parts to this research problem. First, we need a broad array of metrics that reflect various instances of our partial network adversary. Second, we need to understand for each metric how stable it is (sensitivity analysis) and what changes in the Tor network would efficiently improve (or harm) it. Third, we should make the calculations as realistic as possible based on actual Tor client behavior. Last, we need a better infrastructure for comparing the safety impact from alternate design scenarios, for example different path selection or route balancing algorithms.

We'd love to work with you[2] to help answer these questions.

[1] https://blog.torproject.org/files/relays-diff-2011-01-28.png
[2] https://www.torproject.org/research

1 Part one: new metrics

The network graphs[3] on our metrics site show the number of relays and advertised capacity over time. More importantly, we've archived all the relay descriptors and network consensus documents[4] going back to 2004. So now we can ask some questions to better capture the safety of the network over time.

What is the entropy of path selection over time? When we add really fast relays, the capacity of the network goes up but the safety of the network can go down because it gets less balanced. We can start with a simple approximation of a Tor client's path selection behavior—e.g., choose the first hop randomly weighted by bandwidth, and choose the final hop randomly weighted by bandwidth from among the nodes that can exit to port 80. I'm particularly interested in the entropy of the first hop and the third hop, since those are the key points for the anonymity Tor provides against a partial network adversary. Calculating entropy on the probability distribution of possible first hops is easy (as is last hops), and it's worth looking at these individual components so we have better intuition about where our anonymity comes from. Then we'd also want to look at the unified metric (probability distribution of first cross last), which in the simple approximation is the sum of the two entropies.

Then we should do the same entropy calculations, but lumping relays together based on various shared characteristics. For example, country diversity: what's the entropy of path selection over time when you treat all relays in Germany as one big relay, all relays in France as one big relay, etc? How about autonomous system (AS)[5] based diversity, where all the relays in AT&T's network are lumped together?

Another angle to evaluate is what expected fraction of paths have the first and last hop in the same country, or the same AS, or the same city, or the same continent. What expected fraction of paths cross at least one ocean? How about looking at the AS-level paths *between* relays, like the research from Feamster [2] or Edman [1]? What user locations are more safe or less safe in the above metrics?

It's plausible to imagine we can also gain some intuition when looking at the *possible diversity* rather than the entropy. How many ASes or countries total are represented in the network at a given time?

The goal here isn't to come up with the one true metric for summarizing Tor's safety. Rather, we need to recognize that there are many threat models to consider at once, so we need many different views into what types of safety the network can offer.

2 Part two: how robust are these metrics?

For each of the above metrics, how stable are they when you add or remove a few relays? Are certain relays critical to the safety of the Tor network? One way to look at this question is to graph the fall-off of safety as you remove relays from the network, in one case choosing victims randomly and in another choosing them to optimally harm the network. In the latter case, I expect it will look pretty dire. Then look at the same scenario, but now look at the safety of the

[3] https://metrics.torproject.org/network.html

[4] https://metrics.torproject.org/data.html#relaydesc

[5] http://en.wikipedia.org/wiki/Autonomous_system_%28Internet%29

network as you remove *capacity* from the network. For example, consider that the adversary's cost of removing a relay is equal to its capacity. Then explore these attacks again, but rather than looking at attacking individual relays look at attacks on ASes, countries, or continents.

Then look at the robustness over time: is an adversary that can knock out X% of the capacity in the network (for various values of X) getting more or less influential as the network grows? How about an adversary who can knock out an absolute capacity of K bytes for various values of K?

From the other direction, are there certain geographic or network locations where adding relays with a given capacity will most influence your safety metrics? This influence could be positive ("where should I put my relay to best help the Tor network?") or negative ("where should I put my relay to best attack users?").

Is there any relationship between the rate of new relay arrival (e.g. during political events or after popular conferences) and the level of diversity of these relays?

3 Part three: how good was the simple approximation?

While the above calculations assume a simplified path selection model, the reality[6] is more complex. Clients avoid using more than one relay from a given "/16" network or family[7] in their paths. They only use relays with the Guard flag[8,9] for their first hop. They read weighting parameters from the consensus and use them to avoid Guard-flagged nodes for positions other than the first hop and avoid Exit-flagged nodes for positions other than the last hop, in proportion to how much capacity is available in each category. They track how long it takes them to build circuits, and preemptively discard the slowest 20% of their paths.

Accounting for all of these behaviors (especially the last one) will be hard, and you'll want to work closely with the Tor developers to make sure you're moving in the right direction. But even if you don't handle all of them, having a more realistic sense of how clients behave should allow you to better capture the safety in the live Tor network. If you want extra feedback, you can compare your path selection predictions with paths that Tor chooses in practice[10] (full data here[11]).

We should rerun the above experiments with the more accurate client behavior, and see if any of the results change substantially, and if so, why. These changes are great places to recognize and reconsider tradeoffs between what we should do for maximum safety according to these metrics vs what we should do to optimize other metrics like performance and vulnerability to other attacks. Some of these differences are intentional; for example, we don't give the Guard flag to every relay because we want to reduce the churn of entry guards. But others are due to design mistakes; for example, we are probably not giving the Guard flag out as broadly as we should, and I imagine that really hurts the network's safety.

How far away is the "optimal" approximation curve you produced in part one from the "in

[6] https://gitweb.torproject.org/torspec.git/blob/HEAD:/path-spec.txt

[7] https://www.torproject.org/docs/faq#ManyRelays

[8] https://www.torproject.org/docs/faq#EntryGuards

[9] https://gitweb.torproject.org/torspec.git/blob/HEAD:/dir-spec.txt

[10] https://metrics.torproject.org/data/moria-50kb.extradata

[11] https://metrics.torproject.org/data.html#performance

practice" curve here? How does the divergence from the optimal look over time? That is, were we at 50% of optimal back in 2007, but now we're only at 20%?

4 Part four: consider alternate designs

Once we've looked at how safe the Tor network has been over time for our broad array of metrics, we should do experiments to evaluate alternate network designs.

Examples include:

A) If we give out Guard flags according to some other algorithm, how much safety can we get back?

B) In Tor 0.2.1.x we started load balancing based on active bandwidth measurements,[12] so we use the bandwidth weights in the network consensus rather than the weights in each relay descriptor. We know that change improved performance,[13] but at what cost to safety?

C) What happens to the network diversity if we put a low cap[14] on the bandwidth weighting of any node that hasn't been assigned a measurement by the bandwidth authorities yet, to protect against relays that lie about their bandwidth? If there's no real effect, we should do it; but if there's a large effect, we should find a better plan.

D) If we move forward with our plans to discard all relays under a given bandwidth capacity,[15] how will various choices of bandwidth threshold impact the network's safety?

E) How about if we discard the X% of paths with the highest expected latency, as suggested by Stephen Rollyson?[16]

F) What if some Tor users choose their paths to optimize a different network property (like latency or jitter), as suggested by Micah Sherr [3]? The infrastructure you've built to measure diversity here should apply there too.

G) If only a given fraction of exit relays support IPv6, how much does it reduce your anonymity to be going to an IPv6-only destination?

We've put a lot of effort in the past year into understanding how to improve Tor's performance,[17] but many of these design changes involve trading off safety for performance, and we still have very little understanding about how much safety Tor offers in the first place. Please help!

[12] https://gitweb.torproject.org/torflow.git/blob/HEAD:/NetworkScanners/BwAuthority/README.BwAuthorities

[13] https://metrics.torproject.org/performance.html?graph=torperf&start=2009-08-01&end=2009-10-01#torperf

[14] https://trac.torproject.org/projects/tor/ticket/2286

[15] https://trac.torproject.org/projects/tor/ticket/1854

[16] http://swiki.cc.gatech.edu:8080/ugResearch/uploads/7/ImprovingTor.pdf

[17] https://blog.torproject.org/blog/why-tor-is-slow

References

[1] Matthew Edman and Paul F. Syverson. AS-awareness in Tor path selection. In Ehab Al-Shaer, Somesh Jha, and Angelos D. Keromytis, editors, *Proceedings of the 2009 ACM Conference on Computer and Communications Security, CCS 2009, Chicago, Illinois, USA, November 9-13, 2009*, pages 380–389. ACM, 2009. http://freehaven.net/anonbib/#DBLP:conf:ccs:EdmanS09.

[2] Nick Feamster and Roger Dingledine. Location diversity in anonymity networks. In *Proceedings of the Workshop on Privacy in the Electronic Society (WPES 2004)*, Washington, DC, USA, October 2004. http://freehaven.net/anonbib/#feamster:wpes2004.

[3] Micah Sherr, Matt Blaze, and Boon Thau Loo. Scalable link-based relay selection for anonymous routing. In Ian Goldberg and Mikhail J. Atallah, editors, *Proceedings of Privacy Enhancing Technologies, 9th International Symposium, PETS 2009, Seattle, WA, USA, August 5-7, 2009*, volume 5672 of *Lecture Notes in Computer Science*, pages 73–93. Springer, 2009. http://freehaven.net/anonbib/#DBLP:conf:pet:SherrBL09.