Evaluation of User Data Privacy Within Tor

Sean Meek, Ivan Rodriguez Holguin, and Sanchari Das
Department of Computer Science and Engineering
University of Denver
{Sean.Meek,Ivan.Rodriguezholguin,Sanchari.Das}@du.edu

***Abstract*—The Tor network is a popular privacy-preserving protocol often perceived as a one-stop shop for anonymous web browsing. Despite these perceptions, Tor comes with complexity in configuration, security weaknesses, and the potential for de-anonymization. In this work, we simulate network traffic across a closed-circuit laboratory Tor network. Next, we inspect live requests and responses to and from both a Tor browser and a Tor hidden service. Finally, from an operational perspective, we survey the current Tor service landscape for opportunities where users may unintentionally de-anonymize themselves via self-provided data. We conclude by recommending general safe practices for users of Tor and improvements that Tor designers and maintainers could employ further to protect the security and privacy of its users.**

Overview

Initially designed in the 1990s and refined in 2004 by the U.S. Navy, mathematician Paul Syverson, and computer scientists Michael G. Reed and David Goldschlag [1], The Onion Router (Tor) routes traffic through relays, each providing a layer of encryption masking the application data and identities in the network except those immediately adjacent [2]. While its original intent was to shield U.S. intelligence communications online [3], it has since become widespread and is used worldwide [4]. We explored the scenarios in two categories, in which users may be de-anonymized. The first scenario is at the technical level, where we established two Tor laboratories. The first is a ”closed” simulated Tor network and generated test traffic between the nodes. The second is an ”open” live-connected Tor test environment where we sent Tor traffic from a browser out to the Tor network and back to our test host that is serving a static web page as a Tor hidden service.

We tested, monitored, and collected data on Tor network traffic for both requests and responses in each of these cases. Our second explored scenario was how users may de-anonymize themselves when using Tor sites. We selected ten commonly used Tor sites and surveyed the types of data users could voluntarily submit and involuntarily leak to the site. We defined several categories for user data collection: account configuration, settings, open selection (i.e. making a choice from a provided list), open form (e.g., writing a post on a forum), and personally identifiable. Next, we assessed a risk score for each data collection opportunity. We then assess each Tor service with an averaged risk score and, finally, assess the Tor services collectively with an overall averaged risk score.
the Tor protocol has no guarantees of encrypted application

We were surprised to see the egregious lack of HTTPS enforcement in Tor hidden services. HTTPS encrypts the data of web traffic between a client and a server. Unfortunately, data once it leaves the Tor network. Such results indicate that many protections granted by the Tor network are essentially nullified if the application data is not encrypted. Every site surveyed accepted traffic over HTTP and did not ”upgrade” or redirect (automatically or with a visual aid) the user to HTTPS. Moreover, the posted .onion (Tor hidden service URL top-level domain) sites generally found on Clear Websites were displayed and hyperlinked in their HTTP form, absent of any information on what it means for the user to browse over HTTP instead of HTTPS.

Overall, we propose a guide that users can employ to further protect themselves while browsing across Tor. We find that a privacy-conscious Tor user should:

- learn to properly configure and connect to the network
- be aware of the relay guard the user is connected to
- always use certified channels for the transactions
- prevent JavaScript execution in the browser and use an ad-blocker
- always remain alert regarding the intentionally submitted data and unintentionally data leaked
- take the necessary technical precautions to prevent divulging information about or connecting banking and cryptocurrency accounts
- never use usernames, words, or phrases that can be potentially linked to other personally-identifiable accounts

Acknowledgment

We would like to acknowledge the Inclusive Security and Privacy-focused Innovative Research in Information Technology (InSPIRIT) Lab at the University of Denver. Any opinions, findings, conclusions, or recommendations expressed in this material are solely those of the authors.

References

[1] D. Goldschlag, M. Reed, and P. Syverson, “Onion routing,” *Commun. ACM*, vol. 42, no. 2, p. 39–41, feb 1999. [Online]. Available: https://doi.org/10.1145/293411.293443

[2] A. Macrina and E. Phetteplace, “The tor browser and intellectual freedom in the digital age,” *Reference & User Services Quarterly*, vol. 54, no. 4, pp. 17–20, 2015. [Online]. Available: https://www.jstor.org/stable/refuseserq.54.4.17

[3] R. Dingledine, N. Mathewson, and P. Syverson, “Tor: The second-generation onion router,” Naval Research Lab Washington DC, Tech. Rep., 2004.

[4] M. Edman and P. Syverson, “As-awareness in tor path selection,” in *Proceedings of the 16th ACM Conference on Computer and Communications Security*, ser. CCS ’09. New York, NY, USA: Association for Computing Machinery, 2009, p. 380–389. [Online]. Available: https://doi.org/10.1145/1653662.1653708