Boosting open-world website fingerprinting attacks via outlier exposure and cross-domain adversarial training

Abstract

Website fingerprinting (WF) enables a passive adversary to identify the visited websites by users over anonymity networks. Though WF attacks have been proven effective in closed-world settings, their open-world applicability remains heavily unexplored. Our study reveals that existing SOTA WF attacks become rather imprecise in realistic open-world scenario as they uniformly label the unmonitored website traces from diverse URLs and force the model to fit these heterogeneous samples within the same class. As a result, WF attacks fail to effectively distinguish unmonitored websites from the monitored ones and precisely identify the visited monitored websites. Even worse, the attack performance further degrades under various defenses as the discriminative feature spaces of websites are obfuscated. In this paper, we propose xxx, an efficient training scheme for boosting WF attacks in realistic open-world scenario. Specifically, based on the fact that samples from unmonitored websites are essentially OOD data to those from the monitored websites, xxx incorporate a KL-divergence loss term to maximize model's uncertainty on the unlabeled training samples from unmonitored websites while minimizing the original training loss, e.g., CrossEntropy loss, on labeled training samples from the monitored websites. In the presence of traffic defense, xxx utilizes a cross-domain contrastive learning method to extract the invariant correlation of defended and raw traffic in a defense-agnostic manner, which encourages the attack model from simultaneously distinguishing unmonitored websites from monitored ones, as well as obfuscated traffic from different monitored websites. We combine xxx with four SOTA WF attacks and conduct extensive experiments under challenging open-world settings and seven traffic defenses. xxx improves the r-precision score of SOTA WF attacks by an average of yyy. Furthermore, we collect the traffic of visiting websites through Tor and massive real-world traffic to form an extremely low base rate (1e-4) scenario, where xxx successfully boosts the SOTA WF attacks with an average r-precision improvement of zzz.

CCS Concepts

• **Networks** → **Network privacy and anonymity**.

Keywords

Tor; privacy; website fingerprinting; out-of-distribution detection; contrastive learning

ACM Reference Format:

. 2024. Boosting open-world website fingerprinting attacks via outlier exposure and cross-domain adversarial training. In *Proceedings of the 2024 ACM SIGSAC Conference on Computer and Communications Security (CCS '24), October 14–18, 2024, Salt Lake City, UT, USA.* ACM, New York, NY, USA, 1 page. https://doi.org/10.1145/3658644.3670272

1 Website fingerprinting in the open-world

References

CCS '24, October 14–18, 2024, Salt Lake City, UT, USA

ACM ISBN 979-8-4007-0636-3/24/10

https://doi.org/10.1145/3658644.3670272